HOMOGENEOUS TO HETEROGENEOUS FACE RECOGNITION

MUHAMMAD KHURRAM SHAIKH

PhD

2015

Homogeneous to Heterogeneous Face Recognition

Muhammad Khurram Shaikh

A thesis submitted in partial fulfilment
of the requirements of the
University of Northumbria at Newcastle
for the degree of
Doctor of Philosophy

Research undertaken in the
Faculty of Engineering and Environment

November 2015

Declaration

I declare that the work contained in this thesis has not been submitted for any other award and that it is all my own work. I also confirm that this work fully acknowledges opinions, ideas and contributions from the work of others.

Any ethical clearance for the research presented in this thesis has been approved.

I declare that the Word Count of this Thesis is approximately 29,562 words.

Name: **Muhammad Khurram Shaikh**

Signature: **MKShaikh**

Date: **30th November, 2015**

Homogeneous to Heterogeneous Face Recognition

by

Muhammad Khurram Shaikh

Abstract

Face Recognition, a very challenging research area, is being studied for almost more than a decade to solve variety of problems associated with it e.g. PIE (pose, expression and illumination), occlusion, gesture, aging etc. Most of the time, these problems are considered in situations when images are captured from same sensors / cameras / modalities. The methods in this domain are termed as homogeneous face recognition. In reality face images are being captured from alternate modalities also e.g. near infrared (NIR), thermal, sketch, digital (high resolution), web-cam (low resolution) which further alleviates face recognition problem. So, matching faces from different modalities are categorized as heterogeneous face recognition (HFR). This dissertation has major contributions in heterogeneous face recognition as well as its homogeneous counterpart.

The first contribution is related to multi-scale LBP, Sequential forward search and KCRC-RLS method. Multi-scale approaches results in high dimensional feature vectors that increases computational cost of the proposed approach and overtraining problem. Sequential forward approach is adopted to analyze the effect of multi-scale. This study brings an interesting facts about the merging of features of individual scale that it results in significant reduction of the variance of recognition rates among individual scales.

In second contribution, I extend the efficacy of PLDA to heterogeneous face recognition. Due to its probabilistic nature, information from different modalities can easily be combined and priors can be applied over the possible matching. To the best of author's knowledge, this is first study that aims to apply PLDA for intermodality face recognition.

The third contribution is about solving small sample size problem in HFR scenarios by using intensity based features. Bagging based TFA method is proposed to exhaustively test face databases in cross validation environment with leave one out strategy to report fair and comparable results.

The fourth contribution is about the module which can identify the modality types is missing in face recognition pipeline. The identification of the modalities in heterogeneous face recognition is required to assist automation in HFR methods.

The fifth contribution is an extension of PLDA used in my second contribuiton. Bagging based probabilistic linear discriminant analysis is proposed to tackle problem of biased results as it uses overlapping train and test sets. Histogram of gradient descriptors (HOG) are applied and recognition rates using this method outperform all the state-of-the-art methods with only HOG features.

Thesis Supervisor: Dr. Muhammad Atif Tahir
Title: Senior Lecturer

Thesis Supervisor: Prof. Ahmed Bouridane
Title: Professor

Acknowledgments

I would like to thank Professor Ahmed Bouridane and Dr. Muhammad Atif Tahir for their excellent supervision and patience throughout my PhD. I also would like to thank both of them because of their perseverance editing all my papers and thesis. Thanks to all friends for their support which I have made in Department of Computer Science and Digital Technologies, Northumbria University during my PhD.

My PhD studies are never going to be easy without the endless support of my family especially my late father, Mr. Imran Ahmed. My father is always in my memories and my entire career achievements are dedicated to him. He always praised my potential and academic skills at all stages of my educational career. He provided all financial support needed towards my higher education. My mother prayers also play a tremendous role in reaching to this stage. My brothers and sister are also there whenever I needed them. My wife and six kids are there to support and provide endless help to give me momentum needed for PhD studies.

Last but never the least, all this would never have been possible without my faith in ALLAH, to whom I believe gave me the strength, power and knowledge I needed. Thank you.

Contents

List of Figures

List of Tables

Acronyms and Abbreviation

Acronym/Abbreviation	Meaning
CCA	Canonical Correlation Analysis
CCN	Circular Correlation Norm
CCTV	Closed Circuit Television
CMC	Cumulative Match Curve
CRC	Collaborative Representation Classifier
DOG	Difference of Gaussian
FAR	False Acceptance Rate
GAR	Genuine Acceptance Rate
HFB	Heterogenous Face Biometrics
HFR	Heterogenous Face Recognition
HOG	Histogram of Gradient
LIV	Latent Identity Variable
LBP	Local Binary Pattern
LRC	Linear Regression Classifier
KCRC	Kernel Collaborative Representation Classifier
MPN	Modlaity Pattern Noise
MSLBP	Multi Scale Local Binary Pattern
NIR	Near Infra Red
PCA	Principal Component Analysis
PLDA	Probabilistic Linear Discriminant Analysis
PNU	Pixel Non Uniformity
ROC	Receiver Operating Characteristic
SFS	Sequential Forward Search
SPN	Sensor Pattern Noise
SRC	Sparse Representation Classifier
SSS	Small Sample Size
TFA	Tied Factor Analysis
TPR	True Positive Rate
VIS	Visual

Contributions

- **M. K. Shaikh, M. A. Tahir, and A. Bouridane, Modality Identification for heterogeneous face recognition, Springer Journal of Multimedia Tools and Applications (Accepted May, 2016)**

- **M. K. Shaikh, M. A. Tahir, and A. Bouridane, Probabilistic linear discriminant analysis for intermodality face recognition, IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP), pp. 509-513, 2014.**

- **M. K. Shaikh, M. A. Tahir, and A. Bouridane, Tied factor analysis using bagging for heterogeneous face recognition, IEEE 5th European Workshop on Visual Information Processing (EUVIP), pp. 1-6, 2014.**

Chapter 1

Introduction

The need, importance and use of biometric traits are increasing with every day passing. One can find the biometric applications ranging from immigration control in the airports to the military, defence organizations. In biometric identification, person is identified automatically based on physiological characteristics including iris, fingerprint, face, signature, voice etc. Biometric identification has several advantages over old-fashioned methods that involve passwords, tokens or PINS as it removes the need of remembering these difficult combinations of alpha-numeric and special letters. The main problem associated with password based traditional methods is that one can stole or breach password and gain unauthorised access and entry while in the case of biometrics, they are hard to copy, steal or forge [5].

A biometric recognition system extracts features from person biometric trait and recognizes his/her by matching those features with already enrolled, stored templates in biometric database. This is fundamentally a pattern matching approach. The components of biometric systems include a module for data acquisition, module for feature extraction, module for matching extracted features with stored template features and finally, a module for making a decision to authorise or reject the claimed identity [6].

The success of any biometric system depends on following feature's attribute namely a) uniqueness i.e. each individual should have a unique biometric attribute b) universality i.e. most of the people have the same features c) stability i.e. not changes during the time period d) measurability i.e. computable with ease. [6, 7].

Some of hte biometric modalities i.e face, fingerprint, gait, DNA, iris etc. that carry person's information to discriminate from other individuals are presented in the Figure 1-1.

Figure 1-1: Biometric Modalities e.g. face, iris, fingerprint etc. [1]

Although, face biometrics are readily available with the advancement of technology but its applications in real world scenario are not still fully accepted and materialized due to face manifold non-linear behaviour and performance in challenging setup e.g. occlusion, illumination, pose etc. The problems of face recognition are researched on databases which capture the images using same camera / sensor / modality. This type of recognition is categorized as homogeneous face recognition. Face matching problem is further aggravated when capturing device is changed i.e. NIR camera, thermal camera etc. In some of the cases, sketch has to be drawn due to camera unavailability. This results in matching faces in gallery with high resolution with NIR , thermal or sketch images. This category of intermodality face recognition is termed as heterogeneous face recognition.

The remainder of the introduction to this thesis on homogenous to heterogeneous face recognition is organized as follows. Section 1.1 will discuss the selection of face biometrics for the thesis. In Section 1.2 I will provide an overview of homogeneous face recognition. Advantages and overview of probabilistic heterogeneous face recognition will be presented in Section 1.3 and 1.4 and a very important contribution about the missing module in face recognition pipeline will be discussed in Section 1.5. Finally, Section 1.8 will discuss the organization of the remaining chapters of this thesis.

1.1 Why Face Biometrics?

Face biometric has entered into the main realm of the biometric recognition systems due to availability , non-intrusiveness, social acceptability , easy storage and ubiquity. Faces are available everywhere and should be reckon as future biometric as more and more applications are built upon face recognition technology. The applications based on face recognition can be found in counter terrorism, border access control, day care, banks, voter registration etc [8].

The choice of face biometric compared to other biometrics stems from user non-cooperation as all other biometric modalities require user mutual consent to acquire the sample. While in the case of face biometrics, this is not the case. Faces can be captured without user co-operation e.g. from CCTV [9].

Although, face biometric is being researched over two decades but it is still an active research due to challenges offered in face recognition e.g. Pose, Illumination and Expression (PIE) problem, occlusion problem , appearance difference problem. But I still rate face as future biometric due to its obvious benefits e.g. combining face modality in multi-modal systems. [10]

1.2 Homogeneous Face Recognition

The homogeneous FR considers that images in database are acquired from same camera or sensor. Most famous databases like FERET[11] ,ORL [12], AR [13], PIE [14] and Extended Yale B [15] follow the same rule by capturing all face images with same type of camera with different parameter settings to realize or simulate the challenges encountered by face recognition systems in real-world scenario.

Some of the sample images of FERET database is given in Figure 1-2 which reflects frontal and pose samples. ORL face database samples are provided in Figure 1-3 and AR face database samples in Figure 1-4.

Figure 1-2: FERET database with frontal and pose images

Figure 1-3: ORL database with frontal and expression images

The face recognition starts from identifying or segmenting the face region from the image normally termed as face detection then discriminative features which can capture face shape and texture are extracted and finally, identification or verification is carried out either to narrow down the search of the probe face or confirm / reject the claimed identity. Figure 1-5 reflects the process of facial recognition system [16].

The researchers have designed very rich, powerful feature descriptors in homogeneous FR e.g. LBP [17] , Gabor [18, 19, 20], combination of Local Binary Pattern (LBP) and Gabor [21] etc. In contrast to global features like PCA and LDA, these

Figure 1-4: AR database with images having neutral, smile, anger and scream expression along with occlusion images with sunglass and scarf

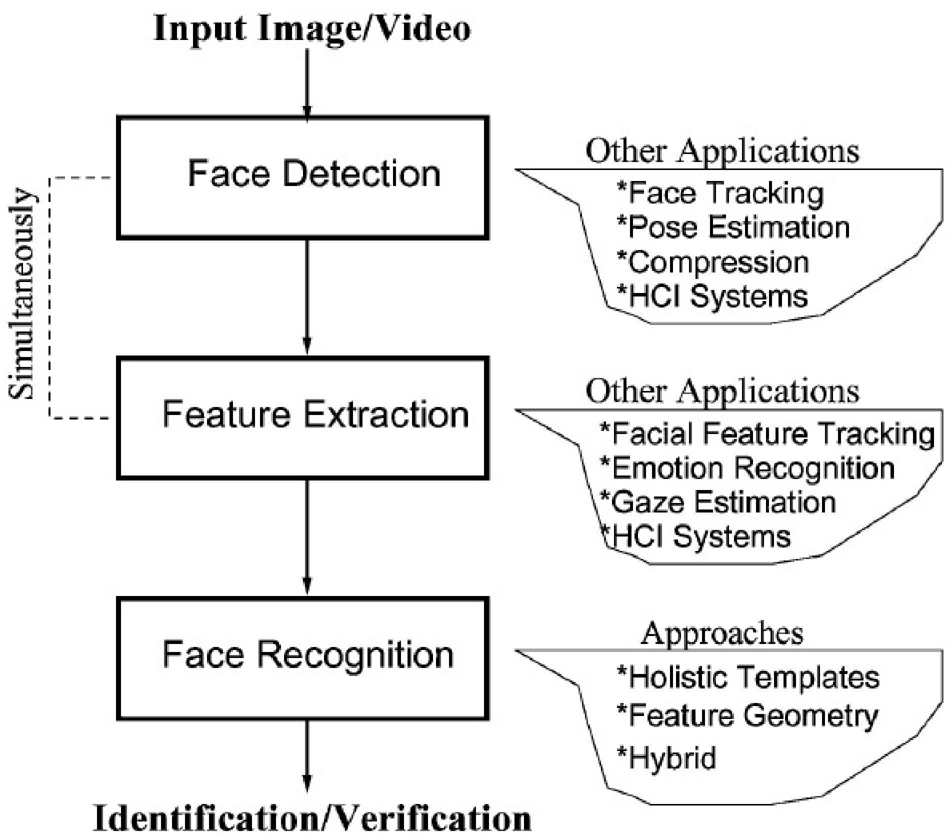

Figure 1-5: Generic Face Recognition System showing face detection , extraction and identification modules

local features possess certain advantages due to their stable performance against expression, occlusion and mis-alignment. Gabor features represent the local structure of the image in relation to their scale and orientation while LBP features give invariance against monotone transformation and somehow display robustness to some extent to illumination changes.

The number on variants of LBP also prove successful in face recognition. Local Ternary Pattern (LTP) [22], Local Quantized Patterns (LPQ) [23] and Multi-Scale Block LBP [24] are LBP-like features.The multi-resolution LBP [25] has resulted in improved performance as compared to single scale LBP. But, the problem associated with LBP features applied at different radii i.e. multi-scale LBP involve high dimensional feature vectors which lead to computationally expensive algorithms. Most recently proposed discriminant face descriptor (DFD, a LBP-like feature) [26] is learned in three steps to extract a discriminant face descriptor. This descriptor has outstanding performance in both homogeneous and heterogeneous face databases.

Although discriminative features play very important role in face recognition success but at the same time, the classifier performance also holds the key to generate higher accuracies. The recent methods which use intensity features and produce some good recognition rates are FR with sparse representation (SRC) [27], Linear Regression for Face Recognition (LRC) [28], two-dimensional PCA (2DPCA)[29] and Collaborative Representation for FR (CRC-RLS) [30]. However, the performance of almost all methods depends exclusively on correct detection, alignment and registration of images. Therefore, in these methods , image normalization plays a very important role in getting higher recognition rates.

1.3 Advantages of Probabilistic Approaches in Face Recognition

Majority of face recognition approaches is based on distance metric; the stored template features are compared with query features with some distance metric e.g. normalized correlation, Euclidean distance etc. and finally, the decision of match or non-match is made [31]. But these distance based methods suffer in the drop of recognition accuracy in the presence of significant difference of extracted face feature vectors of query and template. The reason of this feature variation is attributed to problem of pose changing, illumination changing, modality difference or appearance difference between probe and gallery [32].

PCA and LDA techniques use Euclidean distance metric in frontal face recognition and report good recognition accuracies when there are little image variations exist[33, 34]. Euclidean distance metric is normally considered as hard distance metric because in case of the high degree of uncertainty due to variation in pose, illumination or inter-modality problems, it is not a good choice to assign a hard distance based metric. Soft distance metrics assign a probability to the decision and it will definitely aid in large image perturbations. A probabilistic framework will definitely allow us to defer the decision and collect more samples to reduce uncertainty in reaching the final decision. Another obvious advantage of probabilistic approach is to combine information from other sensors , modalities or image regions with ease.

The other advantages of probabilistic approaches that they can represent the process of image generation including original image and noise termed as generative models. These models reflect the underlying structure which may represent the process of inference. Further, noise component can be modelled carefully [2].

1.4 Probabilistic Heterogenous FR

The probabilistic based homogeneous face recognition is first introduced by B. Moghaddam et. al. [35]. They propose the similarity measure using probability that applies Bayesian analysis on image differences. They learn the likelihoods of intra-personal difference due to expression, lighting and extra-personal variations. The posterior metric is then applied to compute similarity between images. In their earlier work [36],they claimed that " From a probabilistic perspective, the class conditional density $P(x|\Omega)$ is the most important object representation to be learned. This density is the critical component in detection, recognition, prediction, interpolation and general inference". Bayesian Face recognition is another milestone from B. Moghaddam et. al. [37] to announce the arrival of a non-Euclidean similarity measure in face recognition domain and can be viewed as generalized non-linear extension of LDA [33]. C. Liu et. al. [38] proposed a probabilistic approach using Gaussian analysis for modeling the data for each individual. This method only describes the data just after projection and cannot be considered as a fully probabilistic approach. The other major contributions using probabilistic framework for homogeneous face recognition are Probabilistic Linear Discriminant Analysis (PLDA) [2], [39] and Tied Factor Analysis (TFA) [32].

Heterogeneous face recognition brings three major problems namely high intra-class variability , feature gap and appearance difference. The challenge of matching high resolution gallery images with heterogeneous images is evident from the Figure 1-6. To solve these problems in inter-modality matching, biometric researchers handle it by synthesizing one modality samples into another modality to reduce the appearance difference [40], [41], [42]. Other efforts in this domain are to learn and design discriminative features to suppress the effects of feature gaps arisen from VIS-NIR modalities [43], [44], [45]. Researchers in this area have also explored variety of subspace learning approaches to tackle the appearance and feature level differences to VIS-NIR face matching [46], [47], [48]. The details of proposed methods and related works in these three different directions are discussed in Chapter 3, Chapter 4, Chapter 5 and Chapter 6.

But surprisingly, no state-of-the-art methods in heterogeneous face matching until now do any research on probabilistic approach to tackle problems mentioned above. So, I have taken an initiative to apply probabilistic framework getting inspiration from probabilistic based homogeneous face recognition methods. The main contributions in this domain are discussed in detail in Chapter 3 and Chapter 4.

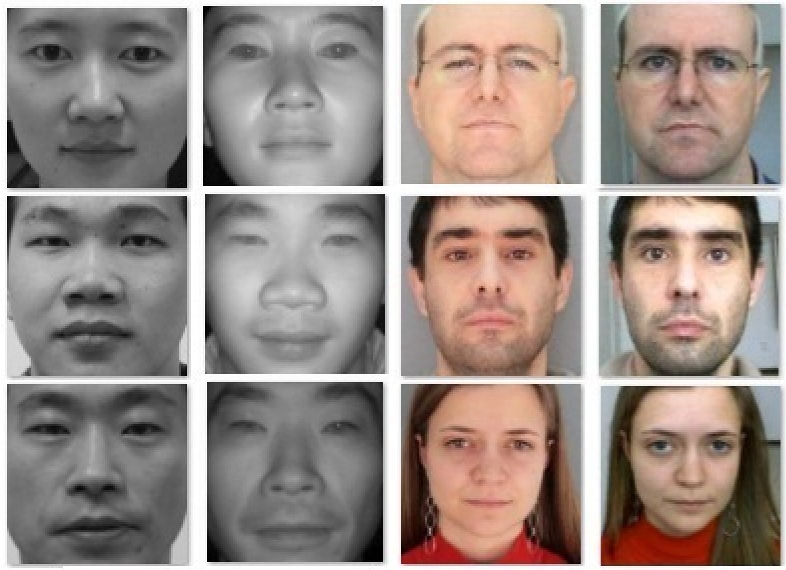

Figure 1-6: 1st and 2nd column: VIS and NIR images HFB database; 3rd and 4th column: Digital and web cameras images Biosecure Database

1.5 Missing Module in HFR

Identifying the type of modalities of the query image which can be of types visual, NIR, digital camera, web camera, etc. have been assumed to be available before face matching. This leads to a major drawback in achieving fully automated heterogeneous face recognition as real world scenarios cannot be reflected. Therefore, modality identification is an important component of the face recognition system which is being overlooked by a majority of the state-of-the-art methods. This component should be given similar attention when comparing with other modules identifying pose, gesture, camera source, etc. In this paper inspired from sensor pattern noise (SPN) estimation based approaches, a novel image sharpening based modality pattern noise technique is proposed for modality identification.

Until now, the research community working on HFR problem has managed to maintain different types of HFR face databases individually [49, 50, 51]. Researchers have applied their proposed methods on each of the database separately and reported the performance accordingly. But with the recent advancement of face recognition technology, a need will arise to identify the modality of face for bringing automation to this field. FR system should be intelligent enough to recognize the modality of the face and forward the request of matching probe face with template in gallery to the correct HFR method handling the identified modality type of the face.

In the future, FR either homogeneous or heterogeneous will move towards automation and will result in deployment of the smart gadgets. The automation of FR technology requires no manual intervention or manual identification of source modalities. To recover modality fingerprint from the source, this thesis proposed unsharp masking based Modality Pattern Noise estimation method.

1.6 Heterogeneous Face Databases

In this thesis, the three challenging heterogeneous face databases have been selected i.e visual vs. NIR (HFB Face Database), low resolution (Web Camera) vs. high resolution (Digital Camera) (Biosecure Face Database)and visual vs sketch (CUHK Face Database).

The HFB [49] is maintained in Center for Biometrics and Security Research and National Laboratory of Pattern Recognition, CASIA contains 202 individuals with total 5097 images out of which 2095 and 3002 images are of visual and NIR modality, respectively. To overcome the illumination problem in face recognition, applications are developed which utilize near infra-red technology based cameras to capture NIR face images and compare them against visual face images in the stored database. The example images from HFB heterogeneous face databases is provided in figure 1-7 containing VIS-NIR samples.

The Biosecure Database [50] contains 420 subjects with 12 samples taken in two sessions. Each session has 6 samples from each individual. Two samples has been captured with webcam while rest with digital camera consisting of flash and non-flash versions in each session. Biosecure database is regarded as multi-scenario and multi-environment database. The example images from Biosecure face face databases is provided in figure 1-8 having digital-cam vs. web-cam images.

CUHK [51] is a publicly available dataset containing 188 subjects of two different modalities i.e face vs. sketch. In training , 88 face-sketch pairs are used while testing

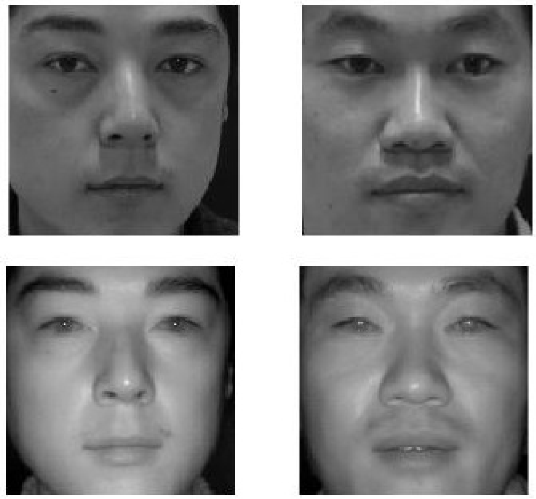

Figure 1-7: HFB database with visual and NIR images in first and second rows respectively.

Figure 1-8: Biosecure database with digital camera and web camera images in first and second rows respectively.

dataset consists of 100 pairs. The example images from CUHK face database is provided in figure 1-9 containing photo-sketch pairs.

Figure 1-9: CUHK database with photo and sketch images in first and second rows respectively.

1.7 Motivations

Multi-scale approaches bring dimensionality curse problem that results in large size of feature vectors and increases computational cost of the proposed approaches. Selection of optimal size of training feature vector is one of the solutions. In Chapter 2, sequential forward approach is adopted to select optimal feature vector. During merging of features of the scale that results in highest recognition rate, this study brings interesting facts that it results in the reduction in the variance of recognition rates among individual scales.

Probabilistic Linear Discriminant Analysis (PLDA) [2] is a complete Bayesian, generative and probabilistic method in homogeneous face recognition domain that produces outstanding results in the presence of extreme pose variations. In Chapter 3, an extension of PLDA for heterogeneous face recognition is proposed that provides a theoretical foundation for this new research area. Due to its probabilistic nature, information from different modalities can easily be combined and priors can be applied over the possible matching. To the best of my knowledge, this is first study that aims to apply PLDA for intermodality face recognition. Also, a very challenging database namely Biosecure database having two different modalities representing real world scenario i.e. web-cam images vs. digital-cam images are used in HFR domain.

The Tied Factor Analysis (TFA) [32] works only with one training pair of same modality i.e. two images from different poses therefore TFA is an specific Bayesian approach in homogeneous face recognition domain that solves single sample space problem. In Chapter 4, TFA is extended to handle heterogeneous modalities i.e. two images from different sensors e.g. VIS-NIR and Webcam-Digcam. Still the results from TFA can be misleading and biased as one can report results on the subset of images which produced high accuracies. So in order to overcome this deficiency of traditional TFA, Bagging based TFA method is proposed to exhaustively test face databases in cross validation environment with leave one out strategy to report fair and comparable results.

The identification of modalities has same similar importance in Biometric systems comparing with pose identification, camera identification, liveliness detection, gesture recognition etc. The modality identification problem is basically ignored in majority of state-of-the-art heterogeneous face recognition methods and all these methods have assumed availability of the image modality information. The module to identify modalities is missing in face recognition pipeline so in chapter 5, modality pattern estimation (MPN) methods are proposed to assist in automation of heterogeneous face recognition approaches by this important component of face recognition.

The new proposed methods inspired from PLDA and TFA are compared in local and global experimental settings with other state-of-the-art methods using Histogram of gradient descriptors (HOG) in chapter 6. Bagging based PLDA is proposed to extensively evaluate heterogeneous face databases to report fair and comparable results. The recognition rates of my proposed methods using only HOG features outperform all state-of-the-art methods and depict the strength and utilization of HOG features application in HFR domain.

1.8 Thesis Organization

In Chapter 2, I propose a kernel version of CRC method using multi-scale LBP based on sequential forward search on the radius parameter to extract most optimal feature subspace. In Chapter 3, I present a probabilistic solution to the problem of heterogeneous face recognition among visual vs. NIR , photo vs. sketch and digital-camera vs. web-camera images. In Chapter 4, a bagging Tied Factor Analysis (TFA) is presented to solve small sample size problem in HFR domain and also remove the biased nature of TFA to report inconsistent results. Chapter 5 brings an important finding on the missing module in face recognition pipeline which is a hindrance in bringing automated heterogeneous face recognition. Chapter 6 studies and evaluates the efficacy of the proposed bagging PLDA and TFA when used in conjunction with HOG features. Finally, I conclude with the findings of this dissertation and suggestions for future research in Chapter 7.

Chapter 2

Multi-scale LBP features selection for Homogeneous Face Recognition

2.1 Introduction

Human Face recognition is a very active research area comprising of variety of face descriptors [52] and classifiers to tackle the problems associated with it namely pose, expression, illumination, occlusion etc. But, due to in built non-linear effects attached with face, it poses a great challenge to researchers for developing a universal technique [16], [53], [54], [55]. High dimensionality, image localization or face detection, stable features and need of strong classifiers are some of great challenges faced by research communities.

As opposed to normal division of FR methods [56] , I consider its division into two broad categories as either intensity or facial-feature based. Intensity based methods use the raw pixels as face features input to the classifier to make the decision. The recent methods which use intensity features and produce some good recognition rates are FR with sparse representation (SRC) [27], Linear Regression for Face Recognition (LRC) [28], two-dimensional PCA (2DPCA)[29] and Collaborative Representation for FR (CRC-RLS) [30]. However, the performance of almost all methods depends exclusively on correct detection, alignment and registration of images. Therefore, in all these methods, image normalization plays a very important role in getting higher recognition rates.

Facial-feature based methods transform the raw pixels into new feature subspace to get more discriminative and strong representation of face to assist in accurate classification of images. Eigenface [34] and FisherFace [33] are one of the prominent facial-feature based methods which revolutionize the research in the area of face recognition. LBP features proposed in [17] are used in lot of FR methods and they normally outperform the Gabor feature based counterparts. The multi-resolution LBP [25] has resulted in improved performance as compared to single scale LBP. But, the problem associated with LBP features applied at different radii i.e. multi-scale LBP involve high dimensional feature vectors which lead to computationally expensive algorithms and curse of dimensionality.

To solve this problem, this chapter tackles the issue of the high dimensionality firstly by down-sampling the images as the feature space choice is no longer critical as reported in [27],[28]. Secondly, I further handle dimensionality issue resulting from multi-scale LBP representation and non-linear behavior of face manifold by projecting the proposed features to kernel subspace.

In [30], Zhang et al. propose an efficient regularized least square classifier based on CRC-RLS for face recognition which outperforms SRC [27] convincingly using l_2 norm. The collaborative representation of a query sample in contrast to sparsity for classification as used in most previous works [57],[58],[59] is applied with low computational burden along with regularized least square method. This approach (CRC-RLS) yields high recognition accuracies. However, this method views the image as a point in a feature space, and thus could not withstand severe illumination alterations. In contrast, histogram-based features, such as the Local Binary Pattern Histogram (LBPH) [60] has gained reputation as a powerful and attractive texture descriptors showing excellent results in the event of extreme lighting problem.

In this chapter inspired from KCRC-RLS method [61], the face image is first represented by multi-scale Local Binary Patterns Histogram (MLBPH) [62] which extends the single-scale representation of LBP features to a multi-scale representation. This multiscale representation is shown to be more accurate than single scale representation of LBP and is also robust to blur and illumination [25].

The main contribution inspired is summarized as follows:

- Multi-scale approaches bring dimensionality curse problem as it results in high dimensional feature vectors that increases computational cost of the proposed approach. Further due to high dimensionality, the algorithm may not generalise owing to small training samples and trap into singularity problem. The other issue with high dimensionality is overtraining problem as it results in complex optimal processing and slow convergence. To overcome these issues, an in-depth study is carried out on the features acquired from individual radii / scales of LBP method. The purpose is to find an optimal feature subspace that accurately represents the training samples and also generalise well with KCRC-RLS method [61]. Sequential forward approach is adopted to analyze the effect of multi-scale. The recognition rates from individually selected radii / scales of LBP are calculated. The radius of LBP which gives highest recognition rate, its features are merged to other radii / scales of LBP features and KCRC-RLS method is applied again to calculate the recognition rates on individual radii. The process is repeated till all the radii features are exhausted. This study brings an interesting facts about the merging of features of individual scale that results in highest recognition rate. After few iterations of merging the individual scale feature with other scales, there is a significant reduction in variance of recognition rates among individual scales. This will provide opportunity to the proposed method to apply any scale low-dimensional feature vector. It will genuinely increase the computational cost and removes other problems associated with high-dimensional feature vector.

This chapter is organized as follows. Section 2.2 reviews Multi-scale Local Binary Patterns (MLBP). Section 2.3 describes the proposed KCRS-RLS based classifier framework for face recognition in the histogram feature space. Section 2.4 provides the effect of merging features from individual scale LBP to other scales and introduces some very interesting observations. Experiment set-up and results are presented in

Section 2.5 and discussed in Section 5.6. Section 2.7 concludes this chapter.

2.2 Multi-scale Local Binary Patterns (MLBP)

The LBP method, shown in Equation 2.1, is a highly discriminative texture operator which extracts occurrences of various patterns in the neighborhood of each pixel in Q dimensional histogram. The histogram features of LBP are robust to translation and rotation of image and invariant to local gray-scale variations. First, the value of the current pixel, u_c, is applied as a threshold to each of its neighbor $u_j(0, 1, ..., Q-1)$ to obtain a binary number. The LBP operation for $R = 8$ is shown in Figure 2-1. The values of neighbors which do not fall in center of the pixel is estimated by interpolation.

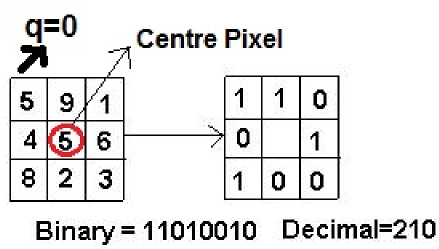

Figure 2-1: Calculation of the LBP.

$$LBP_{Q,R} = \sum_{q=0}^{Q-1} s(u_j - u_c)2^q \tag{2.1}$$

$$s(x) = \begin{cases} 1 & x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

The use of single scale in real world scenarios has shown substantial limitation due to small support area as 3×3 neighborhood will not be able to capture large scale structures which is necessary for acquiring dominant features for some textures. Multiscale LBP has been employed and used successfully [62, 25] to capture structures existed at different scales. The standard LBP operator can be extended to multiscale by using radii of different sizes. Multiscale representation is achieved by sliding set of different radii LBP operators over an image and combining their results to capture non-local information. However, the problem of high-dimensionality is associated with multi-resolution analysis which can be minimized by feature selection technique to curb redundant information. Figure 2-2 provides the examples of different values of Q and R for circular neighborhoods.

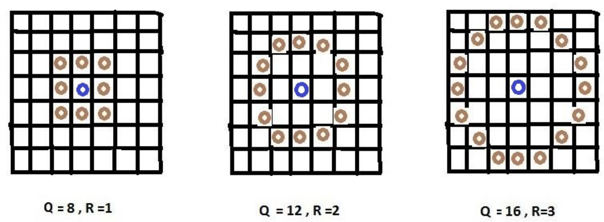

Figure 2-2: Multi-Scale LBP.

By selecting different values of radii depending on the neighbor pixel distance from center pixel, a multiscale representation is created by concatenating LBP histograms from each scale with specific radius value. This multiscale representation is shown to be more accurate than single scale representation of LBP and is also robust to blur and illumination. The face descriptor for multi-resolution analysis is achieved by applying LBP operators at R scales to a face image.The resulting LBP histograms for each scale of image size $M \times N$ is computed by

$$H_Q(l_r) = \sum_{i=0}^{M-1} \sum_{j=0}^{N-1} g(LBP_Q(i,j), l_r), l_r \in [0, L-1] \tag{2.2}$$

$$g(x,y) = \begin{cases} 1 & x = y \\ 0 & \text{otherwise} \end{cases}$$

where $r = 1...R$ and l_r is the maximum bin value of LBP at current r radii.

The computed set of histograms at different radii provides regional information about face image and further concatenated these histograms into a long observation vector. The resultant multi-scale face descriptor is represented by

$$F_{Q,r} = [H_Q(l_1), H_Q(l_2), \cdots, H_Q(l_R)] \tag{2.3}$$

2.3 Kernel Collaborative representation based classification (KCRC-RLS)

CRC-RLS proposed by Zhang et al [30] has established its superiority over its counterparts i.e. SRC [27] and LRC [28] methods. It represents the query sample by exploiting the role of collaboration between classes. The use of l_2 brings favorable characteristics to this method while further stabilising the least square solution using regularization term.

Since, CRC-RLS is a linear method which means its performance will suffer the drop of recognition rates due to existence of non-linear face manifold. The non-linearity in face manifold is caused by variations in illumination, expression, and occlusion. Some of these variations can be addressed by projecting high-dimensional feature vector resulting from multi-scale representation into Kernel space. The non-linearity problem can be handled by CRC-RLS method if the original method is tweaked to handle high dimensional feature vector in Kernel space.

A *kernel* is a function K, such that for all $x_i, x_j \in X$

$$K(x_i, x_j) = < \phi(x_i).\phi(x_j) > \tag{2.4}$$

where ϕ is a mapping from X to an (inner product) feature space F.

Several variants of kernel functions can be applied. In this thesis, I have used RBF kernel with Chi-squared distance metric: $K(x_i, x_j) = e^{-\frac{1}{A} dist_{\chi^2}(x_i, x_j)}$ where A is a scalar which normalizes the distances [63].

The original equation of CRC-RLS [30]

$$\hat{c} = (X^T X + \lambda I)^{-1} X^T y \tag{2.5}$$

can be re-written in kernel space as [61]:

$$\phi(\hat{c}) = (K + \lambda I)^{-1} K_y \tag{2.6}$$

where $(K + \lambda I)$ is positive definite and thus an inverse solution is possible. As equation 2.5 represents the feature vectors of all training class samples using linear approach, the same equation can be extended to handle the high dimensionality of features by using different kernel approaches for subsiding any remaining non-linearities of the manifold.

K_y is a kernel generated from query vector y to all training points. It is clear from Equation 2.6 that $K^* = (K+\lambda I)^{-1}$ is independent of K_y and it can be pre-computed as a projection matrix. The next step is to compute class specific residuals as shown in Equation 2.7.

$$r_i = ||K_y - K^* \times \phi(\hat{c}_i)||_2 / ||\phi(\hat{c}_i)||_2 \tag{2.7}$$

In this thesis, I consider RBF Gaussian kernel for CRC-RLS method. But in reality, there are variety of kernels can be applied e.g. polynomial, spline, ANOVA, Bag of words, tree, graph kernels [64]. Kernel level fusion can be utilized which provides an effective solution to the problem of combining both regional and multiscale face descriptors. This corresponds to taking the Cartesian product of the features spaces of the base kernels (Equation 2.8). Thus, the implicit feature space of the combined kernel is a concatenation of the feature space of the individual kernels. Once the kernels are combined, the proposed KCRC-RLS is used as a classifier.

$$K = \sum_{i=1}^{(r\times R)} K_i \tag{2.8}$$

2.4 Sequential Forward Approach for multiscale LBP histogram features

Calculating LBP histograms on each scale / radius from 1 to R and concatenating all radii features result in one long high dimensional feature vector. This bring high dimensionality problem to KCRC-RLS method and increases the computational cost of this method.

The solution to this problem is to apply feature selection approaches [65] to reduce dimensionality. The feature selection finds optimal feature subspace in lower dimensionality that can accurately represent the actual feature vectors. It can be broadly categorized into filters and wrappers. I am more interested in applying wrapper approach [66]. In this thesis, I calculate the classifier performance on each individual radius / scale and combining the features of that scale which produces highest recognition rates to other scale features. I continue to explore the performance of KCRC-RLC method by combining next stage of sequential forward approach concatenating features of highest recognition performance scale to other scales.

Effect of Sequential Forward Approach to the multi-scale radii LBP features

In this chapter, the experiment on FERET database (Protocol II) in Section 2.5 using KCRC-RLS method with multi-scale LBP features is presented in Figure 2-3. Results definitely authenticate the efficacy of multi-scale approach as recognition rate of single scale is quite lower as compared to multi-scale counterpart. In experiment on FERET database with 400 subjects using protocol II when test sample is "ql", I receive 81.25% rank one recognition rate on single scale i.e. radius = 1 and feature size of 256D. While on multi-scale i.e. radius =8 when all features from previous radii are concatenated to make final feature vector of dimension 2048D, the recognition rate reaches to 93%. The significant rise in rank one rate is approximately 15%. But, the benefit of higher recognition accuracies from multi-scale methods comes at the expense of high dimensionality problem and computational cost. One of the solution to high dimensionality is to find the optimal feature subspace in lower dimension that can accurately represent the data.

Numerous approaches can be found for feature selection [69] but the most popular is sequential forward search. In experiment, I apply the core concept of this approach on my radius parameter to analyze the effect of multi-scale. I calculate the recognition rates on each individual scale using my KCRC-RLS classifier. The scale which provides the highest accuracy, its features are then combined with other remaining scales and I apply this method again to calculate the rank-one recognition rates. The

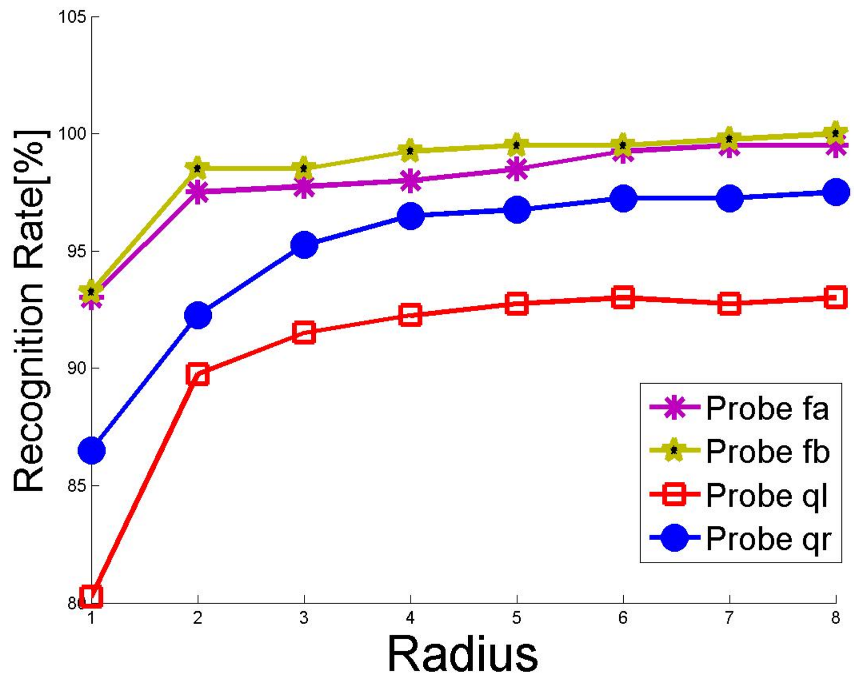

Figure 2-3: Multi-scale LBP with varying radii vs. recognition rate Protocol II FERET.

highest recognition rate radius parameter features are combined again with other scale's features. The process is repeated till all the radii features are exhausted.

During this process of amalgamation of individual scale features with highest recognition rates to other scales features, an interesting observation has been made. Figure 2-4 shows that after applying SFS approach on individual scale when I use radius from 1 to 8, the variance among different radii recognition rates reduces to almost zero. Even though, I display the first four runs of the process of concatenation of scale with highest rate to other scales, the variance reduces significantly. The whole process of feature selection can be applied offline to enhance the performance of biometric system.

It is to be noted that feature selection definitely diminishes the dimensionality problem but may result in lower recognition performance. In this method on the application of full multi-scale features, the recognition performance is 93% with 2048D feature dimension. But after applying SFS approach, the recognition rate is reduced to almost 90% which is still outperforming the comparing methods in Table 2.2 but definitely with substantially reduced lower dimensionality of feature vectors i.e. 256D.

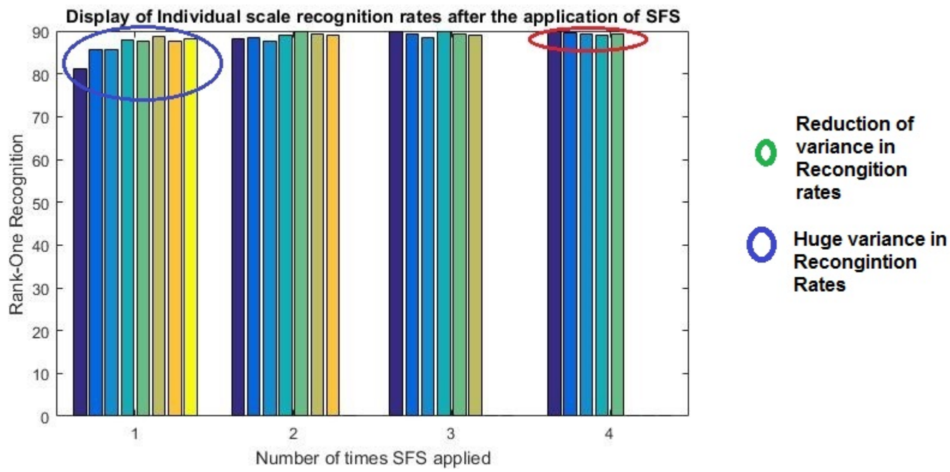

Figure 2-4: Effect on individual recognition rates after the application of SFS approach to FERET database Protocol II using test=ql when radius = 1 to 8.

In order to authenticate my findings, I play around with radius parameter to justify my claim. I select the different radius parameters i.e. 1 to 16, 1 to 32 and 1 to 64 to verify the findings. Figure 2-5 , 2-6 and 2-7 also depict the same finding as after sixth run of SFS, the variance among different scales start diminishing and reduces to almost zero.

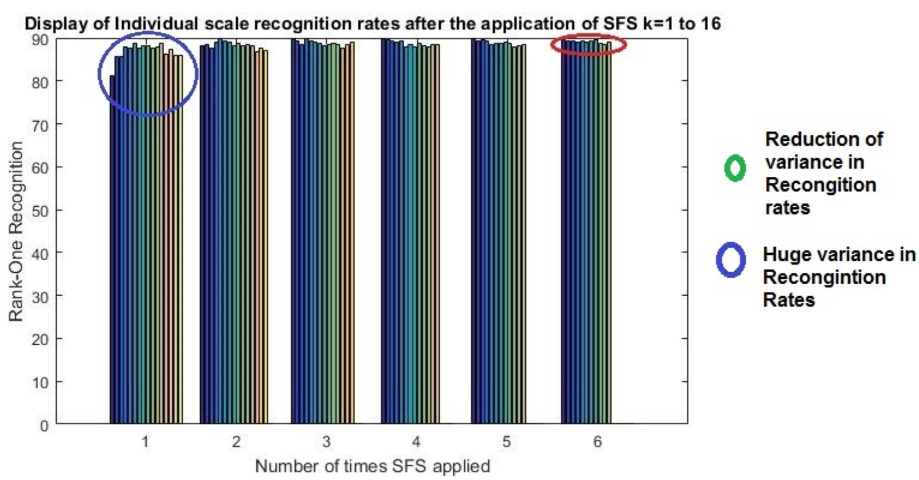

Figure 2-5: Effect on individual recognition rates after the application of SFS approach to FERET database Protocol II using test=ql when radius = 1 to 16.

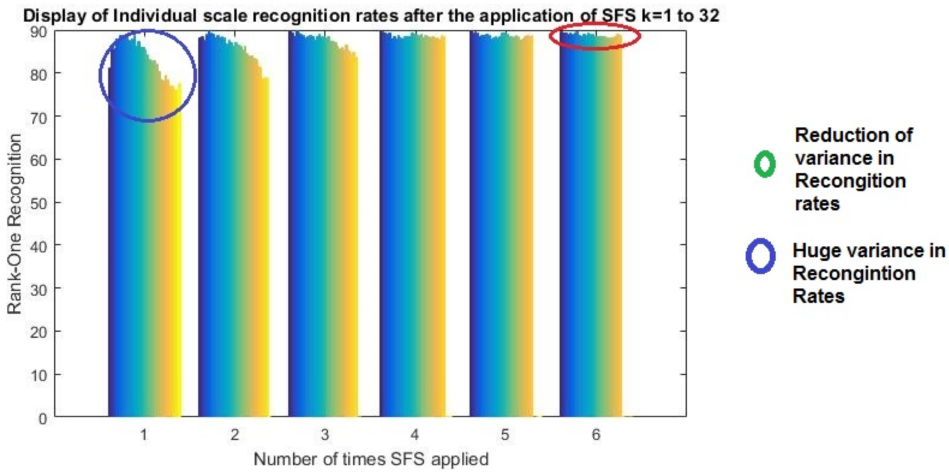

Figure 2-6: Effect on individual recognition rates after the application of SFS approach to FERET database Protocol II using test=ql when radius = 1 to 32.

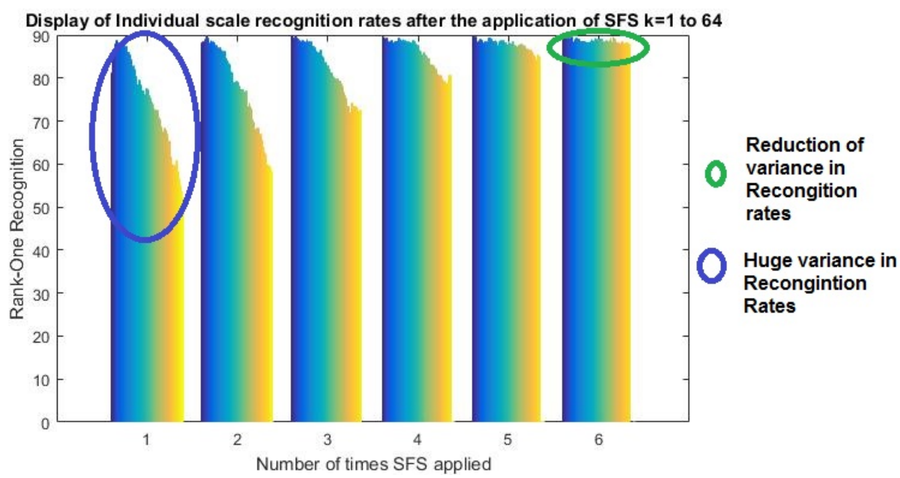

Figure 2-7: Effect on individual recognition rates after the application of SFS approach to FERET database Protocol II using test=ql when radius = 1 to 64.

2.5 Experiments and Results

In order to evaluate the effectiveness of the proposed method, extensive experiments were carried out on four standard databases: FERET[11] ,ORL [12], AR [13] and Extended Yale B [15]. In all experiments, the regularization parameter $\lambda = 0.001$ is used for "face" features as suggested in [30] while λ is set to 0.5 for "histogram" features. It is worth while to mention that I have selected these parameters after extensive simulations.

2.5.1 FERET Database

The FERET database is the largest publicly available database [11] and it is considered as a standard benchmark in reporting recognition rates of major face recognition algorithms.

Protocol I

In my experiment, standard evaluation protocol similar to [28, 67, 68] is used. I use four images of a given subject. Out of four images, galleries are formed containing two frontal "fa" and "fb" and two non-frontal "ql" and "qr" images of each individual. A total of 128 subjects are used in the first protocol. In the training phase, I have used 512 images while the remaining images are used in the testing phase. To achieve reliable collaboration among the classes, more than one training sample is required.

All the experiments are conducted using cross-validation approach in which combination of three samples are used for training while the left over is used for testing. Figure 2-8 shows a typical subject from the FERET database.

Figure 2-8: A typical subject from the FERET database.

For face features, experiments are conducted on down-sampled 7×6 feature subspace. In multi-scale LBP histogram features, an original image is processed with radius varying from 1 to 8 with 8 neighbors. The final dimension of LBP is 2048. Table 2.1 compared the results with state-of-the art methods. However, the face features applied on CRC-RLS approach performs marginally better than PCA, ICA but LRC approach shows improved results when compared to CRC-RLS (Face). The efficacy of using MLBP features in conjunction with RBF kernel version is quite evident as it outperforms the all methods considerably. The increase of 50% and 39% in performance as compared to original CRC-RLS is reported with KCRC-RLS. Furthermore, the results on frontal and non-frontal samples are quite promising as all recognition rates are close to each other except on ql.

Table 2.1: Results for the FERET Evaluation protocol I.

	Recognition Rates (%)			
Method	**fa**	**fa**	**ql**	**qr**
PCA [28]	74.22	73.44	65.63	72.66
ICA [28]	73.44	71.09	65.63	68.15
LRC [28]	91.41	94.50	78.13	84.38
CRC-RLS$_{\mathbf{RAW}}$ [30]	78.91	78.91	46.09	60.16
KCRC-RLS$_{\mathbf{RAW}}$(RBF_{χ^2}) [61]	90.63	92.97	75.00	83.59
KCRC-RLS$_{\mathbf{MLBP}}$($Linear$) [61]	89.84	89.84	71.09	80.47
KCRC-RLS$_{\mathbf{MLBP}}$(RBF_{χ^2}) [61]	**99.22**	**100**	**92.97**	**98.44**

The selection of different radii in multi-scale local binary pattern definitely makes major impact on the recognition rates. The effect of selecting different radius values and its respective recognition rates is shown in Figure 2-9. From the figure, it is evident frontal faces i.e. f_a and f_b have superior recongition rates compared to non-frontal i.e. q_l and q_r faces. Strangely, the recongition rates of the non-frontal face q_r on different radii is quite close or even better to some frontal faces.

Protocol II

The performance measure of any state-of-art algorithm requires testing using large number of samples. In order to validate the consistent performance of my methodology, I have tested KCRC-RLS with 400 subjects of the FERET database selected

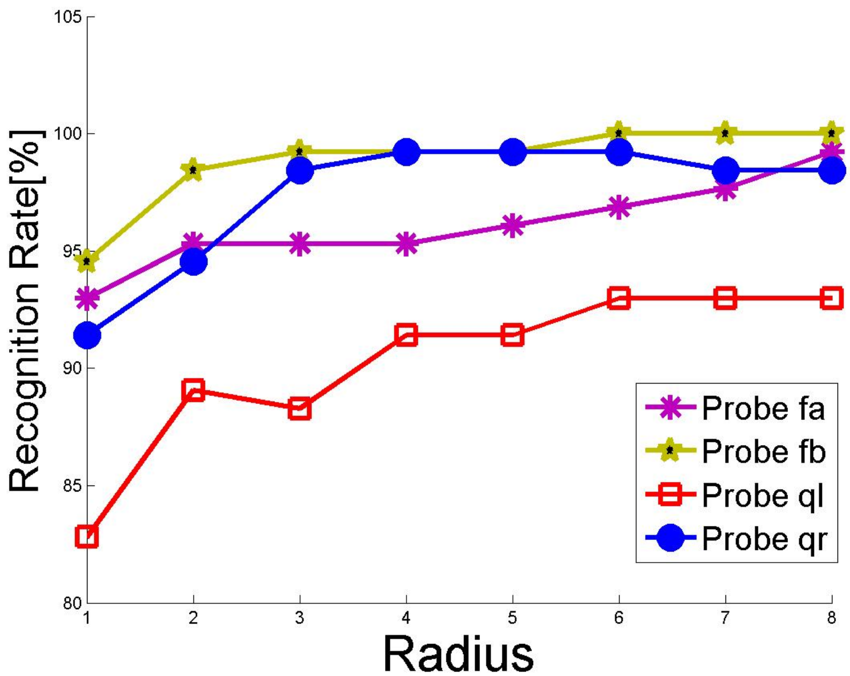

Figure 2-9: Multi-scale LBP with varying radii vs. recognition rate Protocol I FERET.

randomly as compared to 128 subjects in protocol I. I again get consistent results in comparison with protocol I as shown in Table 2.2. The results on frontal samples “fa" and “fb " i.e. 99.50% and 100% and non-frontal samples “ql" and “qr" i.e. 93% and 97.5% are remarkably outstanding and outperforming all reported methods in both protocols.

The effect of selecting different radius values and its respective recognition rates is shown in Figure 2-3.

Table 2.2: Results for the FERET Evaluation protocol II.

	Recognition Rates (%)			
Method	**fa**	**fa**	**ql**	**qr**
PCA [28]	80.00	78.75	67.50	71.75
ICA [28]	93.25	93.50	75.25	76.00
LRC [28]	91.41	94.50	78.13	84.38
CRC-RLS$_{\mathbf{RAW}}$ [30]	71.50	69.25	38.00	39.00
KCRC-RLS$_{\mathbf{RAW}}(RBF_{\chi^2})$ [61]	93.25	92.25	69.50	76.75
KCRC-RLS$_{\mathbf{MLBP}}(Linear)$ [61]	90.00	89.25	72.25	77.75
KCRC-RLS$_{\mathbf{MLBP}}(RBF_{\chi^2})$ [61]	**99.50**	**100**	**93.00**	**97.50**

2.5.2 ORL Database

The ORL database is maintained at the AT&T Laboratories. There are ten different images of each of 40 distinct subjects. For some subjects, the images were taken at different times and under varying lighting. This database also incorporates facial expressions (open/closed eyes, smiling/not smiling) and facial details (glasses/no glasses). Figure 2-10 shows a typical subject from the ORL database.

Figure 2-10: A typical subject from the ORL database.

I follow the standard evaluation protocol given in [29, 28] where the first five images of each individual are used as a training set while the last five are designated as probes. For MLBP, the original image size is processed by 8 MLBP operators with $R = [1, .., 8]$. For this dataset, histogram feature vectors are generated without any non-overlapping region. I have experimented with various non-overlapping regions (2×2, 3×3) and it is observed that the best performance is obtained without dividing the image. As suggested in [28], for $\text{LRC}_{\textbf{Image}}$ and $\text{CRC-RLS}_{\textbf{Image}}$, all experiments are conducted by downsampling 112×92 images to an order of 10×5. Table 2.3 summarizes the results obtained along with a detailed comparison. The table clearly shows the advantage of histogram features in CRC-RLS as there is 4.5% increase in performance when compared with original CRC-RLS classifier i.e. only using "face" features. The performance is further enhanced using the proposed non-linear technique of CRC-RLS with chi-squared RBF kernel. The results have also indicated that the proposed technique also compares favorably with other face recognition methods.

Figure 2-11 shows the relationship of radii in LBP operator with recognition rates. It is clear from graph that multi-scale LBP has brought significant improvement in recognition rates as compared to the uniscale approach.

Table 2.3: Comparison of the proposed technique with other state-of-the-art methods using ORL database.

Method	Recognition Rate (%)
ICA [29]	85.00
Kernel Eigenfaces [29]	94.00
2DPCA [29]	96.00
LRC [28]	93.50
CRC-RLS$_{\mathbf{RAW}}$ [30]	94.50
KCRC-RLS$_{\mathbf{RAW}}(RBF_{\chi^2})$ [61]	95.00
KCRC-RLS$_{\mathbf{MLBP}}(Linear)$ [61]	99.00
KCRC-RLS$_{\mathbf{MLBP}}(RBF_{\chi^2})$ [61]	**100**

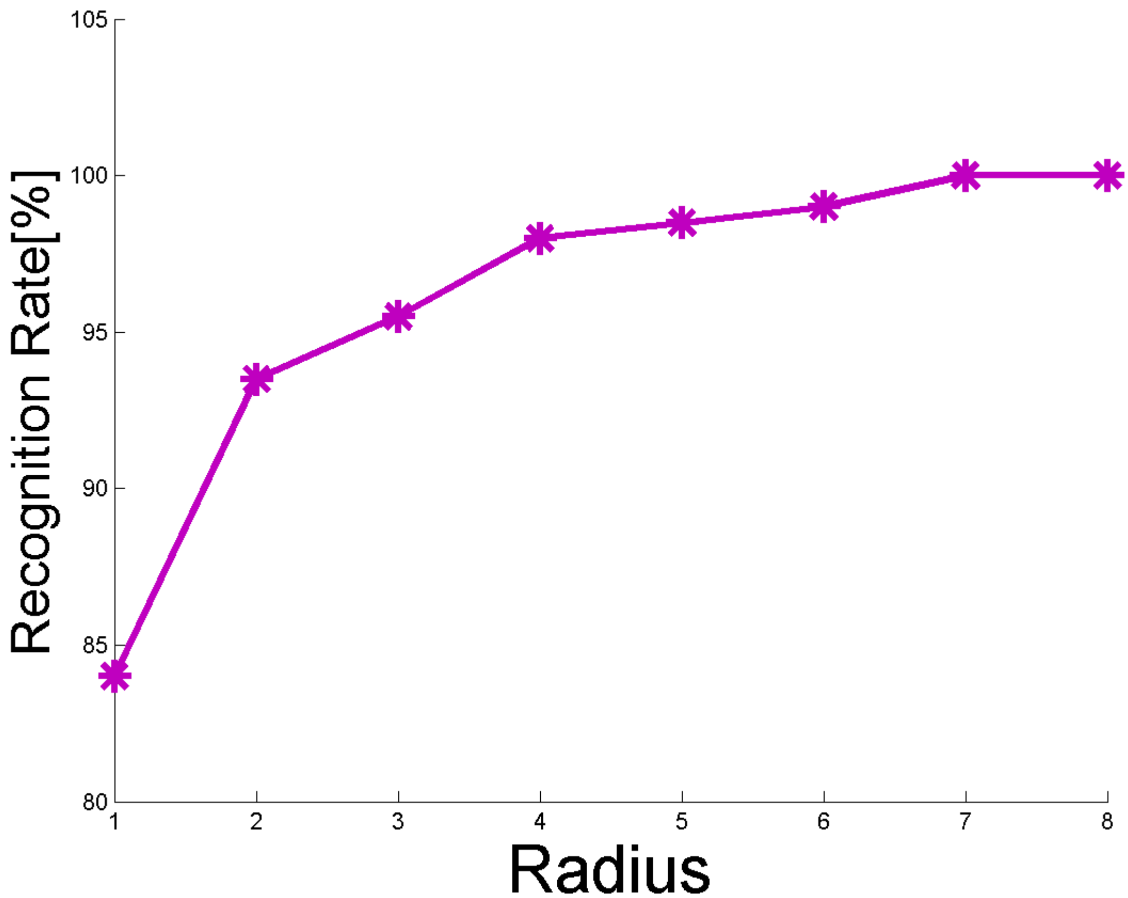

Figure 2-11: Effect of Radius selection on recognition rate.

2.5.3 AR Database

The AR database [13] was developed at the Computer Vision Center at University of Alabama. It contains 4000 images of 126 different peoples out of which 70 are male and 50 are female. This database captures variations in illumination, expression and occlusion. The expression variations consist of neutral, smile, anger and scream while occlusion contains sunglasses and scarf problems.

Experiments related to variations in Gesture

The gesture variation is considered to be one of the challenging tasks in Face Recognition algorithms as it may involve the case where the subject face localization may become difficult due to closure of the eyes. In recent Face recognition techniques, the localization of face or image preprocessing plays an important role for the success of any method dependent on the correct face alignment. The need of methods less dependent on preprocessing stage may definitely subside the risk of getting any stable and practical face recognition methods in future.

I have followed the standard evaluation protocol of [29, 28] as shown in Figure 2-12. In experiment, I have used the same 100 subjects (50 males and 50 females) out of total 128 subjects from AR face database. The gallery for conducting experiments is made up of 100 subjects with four different expressions from two different sessions resulting in total 800 subjects. While, the 600 individuals are used for training and rest for testing. The experiments are carried out again employing the same strategy as in the case of FERET database evaluation. The process is to use three expressions for building the training feature vector while the left over is utilized in making test feature vector.

When conducting experiments on intensity features, all images are downsampled to 10×10 resulting in 100 Dimensional (100D) feature vectors. For MLBP based feature vectors, the same strategy as applied in FERET case is employed with $R = [1, ..., 8]$. The resultant feature vector of MLBP is 2048D. The results are displayed in Table 2.4 showing the KCRC-RLS$_{\mathbf{MLBP}}$ robustness and its superiority over other

Figure 2-12: A typical subjects from the AR database.

reported methods. It clearly outperforms CRC-RLS$_{\mathbf{Face}}$ and LRC by getting 100% recognition rates on all expressions as evident in Table 2.4.

Table 2.4: Results for the AR Gesture Variation Protocol.

	Recognition Rates (%)			
Method	**Neutral**	**Smile**	**Anger**	**Scream**
LRC [28]	99.00	98.50	98.50	99.50
CRC-RLS$_{\mathbf{RAW}}$ [30]	97.50	98.50	98.50	97.50
KCRC-RLS$_{\mathbf{RAW}}(RBF_{\chi^2})$ [61]	98.50	99.00	99.00	99.50
KCRC-RLS$_{\mathbf{MLBP}}(Linear)$ [61]	74.00	67.00	70.50	58.00
KCRC-RLS$_{\mathbf{MLBP}}(RBF_{\chi^2})$[61]	**100**	**100**	**100**	**100**

Experiments related to variations in Occlusion

The challenge of occlusion resulting from the presence of caps, sunglasses, scarves etc poses a major threat for accurate classification of faces. The problems associated with occlusion is to accurately localize the face from image and inaccurate alignment may result in incorrect classification of faces. Furthermore, the consequence of occlusion requires major adjustment for face recognition methods requiring far more dependency on applying manual cropping and alignment of faces.

All experiments do not make use of any pre-processing of images before applying the classification. In this experiment, two different occlusions problems are considered namely: sunglasses and scarfs. Two samples of 100 subjects containing 50 males and females are used. I have built the training feature vector consisting of all samples from previous case i.e., neutral, anger, smile and scream resulting in total of 800 individuals. Secondly, the query feature vectors consist of two sets of occlusion problems as mentioned in Figure 2-13.

Figure 2-13: A typical subjects from the AR database.

The same set of feature vectors for face and MLBP are employed again as in the case of expression variation. Table 2.5 shows the performance using AR Occlusion Protocol. For the sunglasses, KCRC-RLS outperforms the LRC and other methods as I achieve 100% accuracy. This method further outperforms all methods except SRC in the case of very difficult scenario of scarf and getting 50% rise in comparison to LRC and CRC-RLS$_{\mathbf{Face}}$.

Table 2.5: Results for the AR Occlusion Protocol.

	Recognition Rates (%)	
Method	**Sunglasses**	**Scarf**
SRC [28]	87.00	**59.50**
LRC [28]	96.00	26.00
CRC-RLS$_{\mathbf{RAW}}$ [30]	87.00	23.50
KCRC-RLS$_{\mathbf{RAW}}$(RBF_{χ^2}) [61]	96.00	17.00
KCRC-RLS$_{\mathbf{MLBP}}$($Linear$) [61]	97.50	34.00
KCRC-RLS$_{\mathbf{MLBP}}$(RBF_{χ^2}) [61]	**100**	53.00

2.5.4 Extended Yale B Database

This database consists of 2,414 frontal face images of 38 subjects under various lighting conditions. This database is divided into 5 subsets. Subset 1 is used as gallery while others are used for validation (see Figure 2-14). Subset 1 consists of 266 images (7 images per subject) under normal lighting conditions. Subset 2-5 characterize slight-to-severe light variations. Subset 2 and 3 consist of 12 images per subject while subset 4 and 5 consist of 14 images per subject. As suggested in [28], for $\text{LRC}_{\mathbf{RAW}}$ and $\text{CRC-RLS}_{\mathbf{RAW}}$, all experiments are conducted by downsampling to an order of 20×20. For MLBP, the size of input image is 142×120 and the descriptor is generated using 16 MBLP operators with $r = [1, .., 16]$. The image is divided into 4×4 non-overlapping regions for establishing the multi-scale regional histogram.

A detailed comparison of the results is summarized in Table 2.6. It is observed that the original CRC approach ($\text{CRC-RLS}_{\mathbf{RAW}}$) gives excellent performance in moderate light variations and 100% accuracy is observed for subsets 2 and 3. However, the performance falls under severe lighting conditions and 86.5% and 36.9% recognition rates are observed for subsets 4 and 5, respectively. This recognition rate improves significantly for subset 5 when MLBP features are used in CRC-RLS. The accuracy of 86.5% is obtained (an improvement of approximately 57%). This performance is further enhanced using the proposed non-linear solution of CRC-RLS with chi-squared RBF kernel. Results have indicated that there is approximately 15% and 60% improvement for subsets 4 and 5 while using proposed non-linear solution of CRC-RLS. But there is marginal drop of approximately 0.2% in performance for subset 2 when using proposed histogram-based features. The results also indicate that the proposed technique compares favorably with other face recognition methods specially dealing the case of illumination [70].

2.6 Discussion

The results show the usefulness of kernel collaboration at the learning stage for face recognition. The proposed approach successfully captures discriminative informa-

Figure 2-14: Extended Yale B database subjects with various lighting conditions.

Table 2.6: Results for the Extended Yale B database.

	Recognition Rates (%)			
Method	**S2**	**S3**	**S4**	**S5**
PCA [28]	98.5	80.0	15.6	24.4
ICA I [28]	98.0	80.7	17.0	22.0
LRC [28]	**100**	**100**	83.3	33.6
CRC-RLS$_{\mathbf{RAW}}$ [30]	**100**	**100**	86.5	36.9
CRC-RLS$_{\mathbf{MLBP}}$ [30]	99.8	**100**	92.4	81.1
KCRC-RLS$_{\mathbf{MLBP}}(RBF_{\chi^2})$ [61]	99.8	**100**	**97.7**	**93.2**

tion among the different classes by picking up collaboration between classes. This has led to a significant increase in the recognition rates when compared to other state-of-the-art methods. In the case of FERET database (which is mostly used in Face Recognition methods), I have achieved near perfect recognition rates in frontal matching. Figure 2-15 is showing some of the match and non-match cases in the cross identification of four selected scenarios in FERET database considering both protocols.

I further report the efficacy of proposed approach in a very tough and challenging problem of scarf occlusion in AR database. The results on this particular problem is very promising as compared to other methods.

In matching, I use two samples of each subject by creating a test vector in occlusion protocol. It is noticed in my experiment that actually only 30 subjects out of 100 are not classified at all while at least one subject is identified correctly from rest of the subjects. The result on this particular problem can be increased using the approach of partial visible face matching. Some of the faces from scarf occlusion protocol are shown in Figure 2-16 representing correct and incorrect classification. In summary, the presented approach is the best choice among the compared approaches.

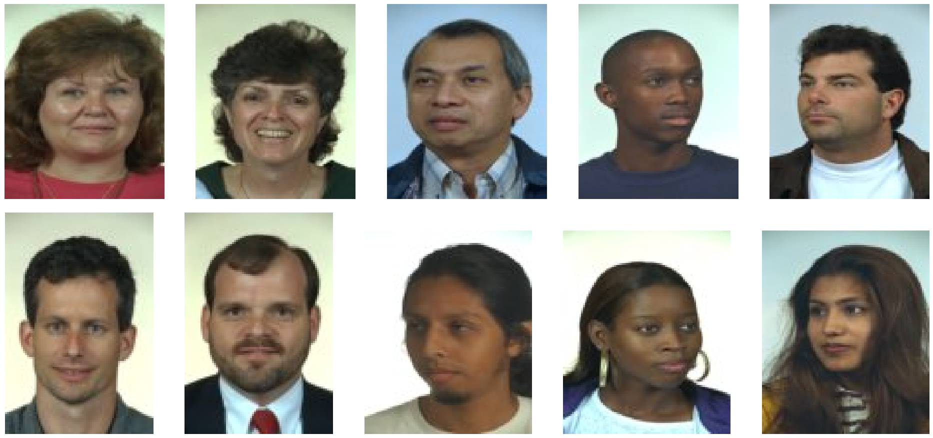

Figure 2-15: First row showing images successfully classified and second row displaying unsuccessfully identified images in both protocols of FERET database.

It provides an efficient and effective learning solution for face recognition.

Figure 2-16: Scarf Occlusion Protocol AR Database : First row showing images successfully classified and second row displaying unsuccessfully identified images

2.7 Summary

In this chapter, I analyse the merging of LBP features from individual radii that has highest recongiition rates when compared among other radii LBP features. Multi-scale approaches bring higher recongition rates at the expense of higher dimensionaltiy of feature vector and computational cost. Finding a optimal feature subspace will tackle these problems. Sequential forward search (SFS) approach finds optimal feature subspace in lower dimensionality that can accurately represent the actual feature vectors. During the merging of features scale that produces highest recognition rate with other scales, an interesting result observed that variance among recognition rates of differnt scales reduces drastically after few iterations of SFS appraoch. Due to this reduction in variance, feature vector of any scale can be used as it is already an optimal feature subspace that produces high recongition rate.

Chapter 3

Heterogeneous Face Recognition : Probabilistic Discriminant Analysis

Face recognition is a difficult problem due to the intrinsic similarity of the classes, the wide range of perturbations and changes in imaging conditions. These include variations in illumination and facial expression, occlusion, and pose or view angle. These challenges are manifested in large variability in facial appearance of the same person. The problem is further aggravated by intermodality face matching involving matching faces from different modalities such as infra-red images, sketch images and low/high resolution visual images.

Recent trends have shown that the researchers in Biometric area are trying to tackle this problem by minimizing the feature gap of the same image captured using different modalities. There are three major and broad categories where researchers can handle this issue. i) Analysis by synthesis methods: face samples of one modality are first transformed to another modality so that the appearance difference is minimized. The representative work in this area include Eigen transform Method [40] ,Local linear preserving Method [41], MRF modeling [42]. ii) Extraction of Consistent Features: proper texture descriptors are designed to reduce the feature gap between modalities. Difference of Gaussian (DOG) filter [43] is used to reduce appearance difference and extract Multi-block LBP. Using HOG and LBP , applying sparse representation classifier [44], SIFT and multiscale LBP [45] is also employed in this area. iii) Subspace learning methods: find a common discriminant subspace to classify heterogeneous data. Some of the representative works in this area are Regularized Discriminative CSR [46] , CDFE [47].

In this chapter, I propose a new method inspired from Probabilistic Linear Discriminant Analysis (PLDA) [2] for heterogeneous face recognition. PLDA is a generative probabilistic method which models the face into signal and noise components. It seeks to maximize the discrimination probabilistically by maximizing the inter-class variation and minimizing the intra-class variance. Further, it is a Bayesian generative approach, thus, brings quite favorable characteristics e.g. allowing careful modeling of noise, ignoring variables of least interest by marginalizing over them and providing a coherent way of comparing models using Bayesian model comparison. Following are the main contributions in this chapter:

- Probabilistic Linear Discriminant Analysis (PLDA) can be regarded as probabilistic equivalent to traditional Linear Discriminant Analysis (LDA) method. This method is providing outstanding performances in homogeneous face recognition domain. In this chapter, I extend the efficacy of PLDA to heterogeneous face recognition. Due to its probabilistic nature, information from different modalities can easily be combined and priors can be applied over the possible matching. To the best of author's knowledge, this is first study that aims to apply PLDA for intermodality face recognition.

- Biosecure face database is considered first time in heterogeneous face recognition research area which deals with web camera and digital camera images. The proposed PLDA method produces outstanding results on this database

The remainder of the chapter is organized as follows. In Section 3.1 , I describe Probabilistic linear discriminant analysis, its training and recognition stages. The experiment set-up and results are presented in Section 3.2. Section 3.3 concludes this chapter.

3.1 Probabilistic Linear Discriminant Analysis(PLDA)

The use of generative probabilistic approaches have been applied in quite wider domain of object recognition [71], image segmentation [72], object tracking [73] etc. The main theme of these approaches lies under the notion that observations are indirectly created from set of underlying variables with some noise associated with it.

PLDA [2] is a generative probabilistic method which models the face into signal component and noise component. The signal component represents the identity of an individual as a hidden variable called as latent identity variable (LIV) while the noise component reflects any remaining variation of the face that is not attributed towards identity.

PLDA is very closely related to Linear Discriminant Analysis [33] as it seeks to maximize the discriminability probabilistically by maximizing the inter-class variation and minimizing the intra-class variance. PLDA is based on Bayesian model, due to this reason it brings quite favorable characteristics e.g. posterior probabilities give more flexility in adjusting/deferring the final decision if uncertainty is quite big.

The other obvious advantage of PLDA as stated in [2] that a probabilistic solution means that I can easily combine information from different measurement modalities and apply priors over the possible matching. For this reason, I extend the utilization of PLDA to intermodality face matching problem.

3.1.1 Latent Identity Subspace (LIV)

PLDA assumes that there exits a multidimensional variable in a new subspace which represents the identity of an individual regardless of the modality. This variable is termed as latent identity variable (LIV) which resides in a subspace called latent identity space as opposed to observed space where the images are captured.

The key property of LIV is that if two LIVs take the same values then it corresponds to an identity of same individual and vice versa. PLDA never measures the LIVs directly but through observed images generated from latent variable with its associated noise. Figure 3-1 reflects latent identity approach.

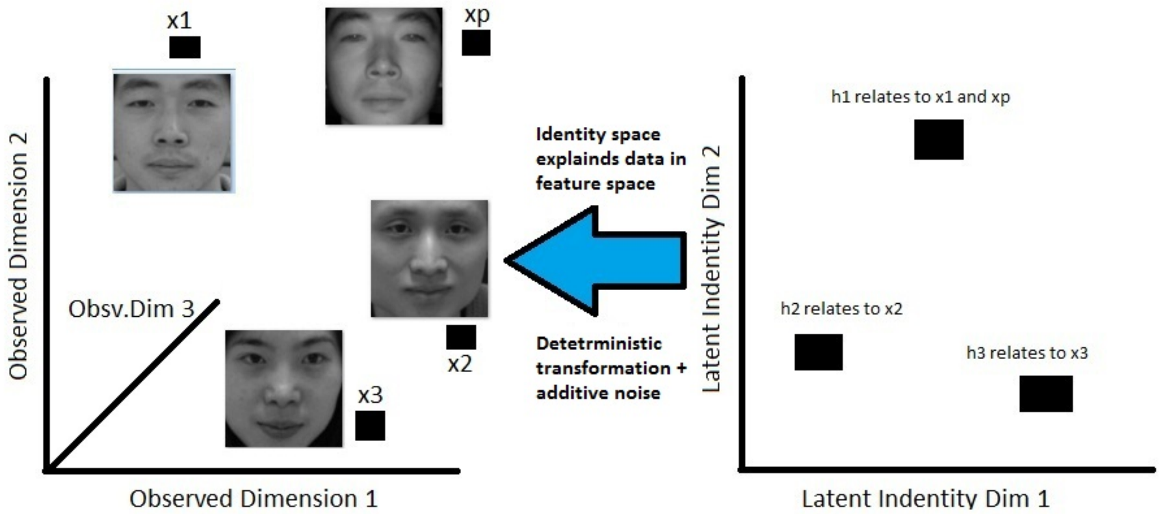

Figure 3-1: Representation of Observed and Identity space showing each point in latent space is different individual while each position in observed space is reflecting different image having been generated from a particular point in latent identity space [2]

3.1.2 PLDA Model Description

PLDA model [2] is of the following form

$$x_{ij} = \mu + Fh_i + Gw_{ij} + \epsilon_{ij} \tag{3.1}$$

It denotes the j^{th} image of the i^{th} individual by x_{ij}. The term μ represents the overall mean of the training dataset. F denotes the basis function for between individual variance with its associated. LIV h_i (remain constant for every person) that corresponds to individual's position in the LIV subspace. G denotes the basis function within individual variance with its associated w_{ij} that corresponds to position in this subspace for j^{th} image of i^{th} individual. ϵ_{ij} is a residual noise term defined as Gaussian with a diagonal covariance Σ.

The signal component of this model $\mu + Fh_i$ depends only on the identity of the person (only i). The reason is that it has no image dependence i.e. no j and it describes between-individual variance. The noise component of the model $Gw_{ij} + \epsilon_{ij}$ depends on both i and j and it describes within-individual variance of the individual with similar images. Formally, the PLDA model can be described using conditional probabilities as

$$P_r(x_{ij}|h_i, w_{ij}, \theta) = g_x[\mu + Fh_i + Gw_{ij}, \Sigma] \tag{3.2}$$

$$P_r(h_i) = g_h[0, I] \tag{3.3}$$

$$P_r(w_{ij}) = g_w[0, I] \tag{3.4}$$

where $g_a[b, C]$ describes a Gaussian in a with mean b and covariance C. Equations 3.3 and 3.4 define simple priors on h_i and w_{ij}. θ is the unknown parameters μ, F, G, Σ.

Figure 3-2 shows the components of PLDA including signal and noise subspace components.

3.1.3 Learning PLDA parameters : Training Stage

In PLDA model, the only known parameters are the observed images while the rest $\theta = \mu, F, G, \Sigma$ are all unknown. If I know h_i and w_{ij} , then the learning parameters F and G will be quite easier. But, unfortunately, all the right hand side parameters of my PLDA model in Equation 3.1 are unknown.

Fortunately, for this chicken-egg problem, one can take advantage of Expectation and Maximization Algorithm [74] which iteratively maximizes the likelihood of parameters alternately in each iteration. The E step finds the unknown identity variables h_i and w_{ij} by calculating posterior probabilities over fixed parameter values. In M step, the algorithm maximizes the lower bound on the parameters $\theta = \mu, F, G, \Sigma$.

Using Equation 3.2, the images from the i_{th} identity $X_i = \{x_{ij}|j = 1, 2, ..., J\}$ form a composite system

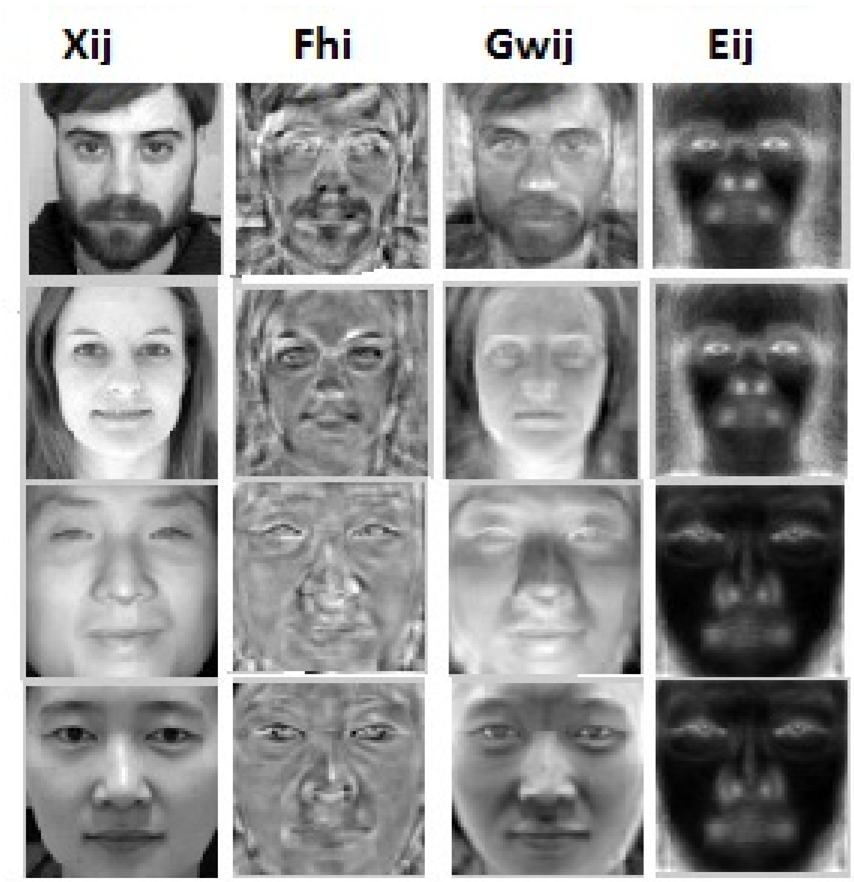

Figure 3-2: Visualization of PLDA signal and noise components

$$\begin{bmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \\ \vdots \\ \mu \end{bmatrix} + \begin{bmatrix} F & G & 0 & \cdots & 0 \\ F & 0 & G & \cdots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ F & 0 & 0 & \cdots & G \end{bmatrix} \begin{bmatrix} h \\ w_1 \\ w_2 \\ \vdots \\ w_N \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_N \end{bmatrix} \tag{3.5}$$

The above formulation can be rewritten as

$$x'_i = \mu' + Ay_i + \epsilon'_i \tag{3.6}$$

The probabilistic form of this composite model is

$$P_r(x'_i|y_i) = g_{x'_i}[\mu' + Ay_i, \Sigma'] \tag{3.7}$$

$$P_r(y_i) = g_{y_i}[0, I] \tag{3.8}$$

The joint distribution of x_i' and y_i is

$$P_r(y_i, x_i'|\Theta) = P_r(x_i'|y_i, \Theta)P_r(y_i) \tag{3.9}$$

The learning consists of two iterative steps: the expectation (E-Step) and maximization (M-Step) procedures.

Expectation (E)-Step: The goal is to estimate a full posterior distribution $P_r(y_i, x'|\Theta)$ of LIVs $y_i = (h_i, w_{ij})$ for each individual separately by fixing the model parameters Θ given the data X.

The first two moments of Expectation (E) steps and update rules for Maximization (M) step for this model [2] are

$$E[(y_i|x_i', \Theta)] = \left(A^T \Sigma'^{-1} A + I\right)^{-1} A^T \Sigma'^{-1}(x_i - \mu') \tag{3.10}$$

$$E[(y_i y_i^T|x_i', \Theta)] = \left(A^T \Sigma'^{-1} A^T + I\right)^{-1} + E[y_i]E[y_i]^T \tag{3.11}$$

Maximization (M)-Step: The goal is to update the values of the parameters $\theta = \mu, F, G, \Sigma$. Setting the joint log–likelihood L_Θ to these parameters to zero respectively and re-arrange to provide the following update rules:

$$\mu = 1/IJ \sum_{i,j} x_{ij} \tag{3.12}$$

$$A = \left(\sum_{i,j} (x_{ij-\mu} E[y_i])^T\right)\left(\sum_{i,j} E[y_i y_i^T]\right)^{-1} \tag{3.13}$$

$$\Sigma = 1/IJ \sum_{i,j} Diag\left[(x_{ij} - \mu)(x_{ij} - \mu)^T - AE[y_i](x_{ij} - \mu)^T\right] \tag{3.14}$$

where $A = \begin{bmatrix} F & G \end{bmatrix}$ and Diag represents only the diagonal elements of the matrix.

The parameters Θ can be initialized randomly. The E-step and M-step are iterated until convergence. The combination of the E-Step and M-Step is guaranteed to increase the overall likelihood of the model at every iteration. The learned parameters can then be used for the following face recognition tasks.

Figure 3-3 reflects the main idea of PLDA that images from different modalities of same subject share the same identity variable.

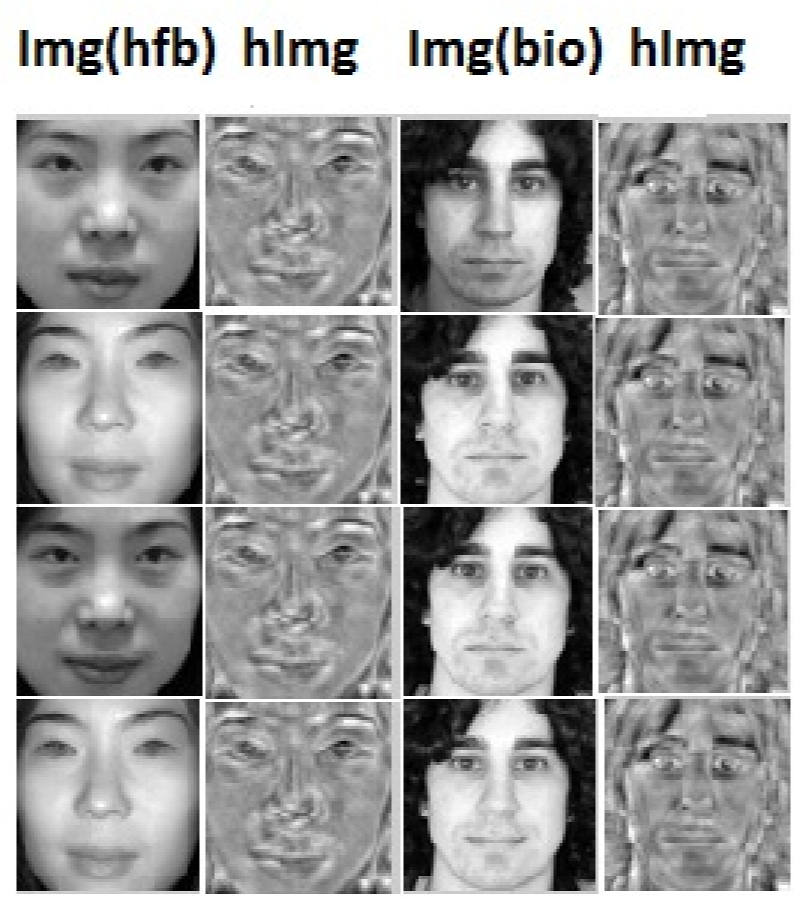

Figure 3-3: The 1st and 3rd columns show images of individual from HFB (VIS-NIR) and Biosecure database and 2nd and 4th columns shows its learned identity variable.It is evident that different modalities images of an individual is represented by same LIV

3.1.4 Recognition Stage

After learning model parameters, the next stage is to match two images sharing the same identity variable h. The recognition stage of PLDA compares the likelihood of the data under N different models which is denoted by $M_{1...N}$. In a closed set identification, the n^{th} model represents the case where probe face x_p matches the n^{th} gallery face so n^{th} identity variable h_n is responsible of generating probe feature vector i.e. $h_p = h_n$. The model M_0 depicts the case where two faces belongs to different people having different identity variables.

The likelihood of observed data $x_i...x_K$ and x_p using Model k can be evaluated as follows

$$P_r(x_{1...K}, x_p|M_k) = \iint P_r(x_1|h_1, w_1)dh_1dw_1 \cdots$$
$$\iiint P_r(x_k, x_p|h_k, w_kw_p)dh_kdw_kdw_p \cdots \iint P_r(x_k|h_k, w_k)dh_kdw_k$$

The evaluation of above integrals is basically the evaluation of likelihood that N images share the same identity variable regardless of noise variables.

For above integrals I have:

$$x_{ij} = \mu + \begin{bmatrix} F & G \end{bmatrix} \begin{bmatrix} h_i \\ w_{ij} \end{bmatrix} + \epsilon_{ij} \tag{3.15}$$

Let $A = \begin{bmatrix} F & G \end{bmatrix}$ and $Z = \begin{bmatrix} h \\ w \end{bmatrix}$ which are the latent variables.

Equation 3.15 becomes a standard factor analyzer: $x = \mu + Az + \epsilon$.

Thus I have

$$P_r(x|z) = g_x[\mu + Az, \Sigma] \tag{3.16}$$

$$P_r(z) = g_z[0, I] \tag{3.17}$$

The likelihood of observing the image assuming that there was no match to any other images can be calculated by marginalizing over the hidden variables z to give:

$$P_r(x) = g_x[\mu, AA^T + \Sigma] \tag{3.18}$$

The second case is when the probe image x_p matches a gallery image x_g. I use the generative equation:

$$\begin{bmatrix} x_p \\ x_g \end{bmatrix} = \begin{bmatrix} \mu \\ \mu \end{bmatrix} + \begin{bmatrix} F & G & 0 \\ F & 0 & G \end{bmatrix} \begin{bmatrix} h \\ w_p \\ w_g \end{bmatrix} + \begin{bmatrix} \epsilon_p \\ \epsilon_g \end{bmatrix} \tag{3.19}$$

or

$x' = \mu' + Bz + \epsilon'$ which again has the form of a standard factor analyzer. The likelihood of the data under this model is:

$$P_r(x_p, x_g) = g_{x'}[\mu', BB^T + \Sigma'] \tag{3.20}$$

where $\Sigma' = \begin{bmatrix} \Sigma & 0 \\ 0 & \Sigma \end{bmatrix}$

Based on Equation 3.19 and 3.20, I can write Equation **??**

$$P_r(x_{1\ldots K}, x_p|M_k) = \frac{\prod_{i=1}^{K} g_{x'_i}[\mu, AA^T + \Sigma] g_{\hat{x}}[\mu', BB^T + \Sigma']}{g_{x'_k}[\mu, AA^T + \Sigma]} \tag{3.21}$$

where $\hat{x} = \begin{bmatrix} x_k \\ x_p \end{bmatrix}$

3.2 Experiments and Results

The robustness of PLDA is tested on two different types of intermodality scenarios i.e Visual vs. NIR, Low resolution (Web Camera) vs. High resolution (Digital Camera). In testing, intensity and LBP features [17] are employed and rank one recognition rates are obtained.

3.2.1 Protocol I and II of HFB VIS-NIR Face Database

In order to test the validity of PLDA in intermodality face matching problem, I employ the same two protocol setups of [46]. The test is carried on VIS-NIR HFB face database [49]. In protocol I, the training set comprises of 1062 VIS and 1487 NIR images of 202 subjects randomly selected while the test set is made up of gallery images from VIS and probe from NIR (the test set is not used in the training set). In protocol II, the training set is made up of 1438 VIS and 1927 NIR of 168 subjects while test set comprises of images from 174 persons using one gallery and probe images not included in the training stage. The samples of VIS-NIR database are of the size 128 x 128 cropped using eye co-ordinates. For intensity and LBP features, all the images are resized to 32 x 32. Table 3.1 compares the recognition rates on HFB using my proposed PLDA and other methods.

Table 3.1: Results for the HFB Evaluation protocol I and II

	Recognition Rates (%)			
Methods	Intensity/PI	LBP/PI	Intensity/PII	LBP/PII
LDA [33]	98.01	98.74	64.51	79.03
CDFE [75]	97.21	99.73	54.87	62.82
LCSR [46]	97.48	99.40	75.65	93.84
LDSR [46]	97.54	99.80	73.96	94.04
KDSR [46]	98.34	99.73	77.04	**95.33**
Proposed PLDA [2]	**99**	**100**	**91.95**	94.25

PLDA reports consistent results on both protocols with intensity and LBP based features. In protocol I and II on intensity and LBP features, it outperforms all the stat-of-art heterogeneous face methods except in PII LBP and reflects its effectiveness over other state-of-art approaches. PLDA validates superiority of the LBP features over other features as it reflects increase in the recognition rates. Cumulative match and Receiver operating curves are also drawn to visualize the performance of proposed PLDA on protocols I and II over HFB database. Figure 3-4 and Figure 3-5 display CMC and ROC curves for protocol I. Figure 3-6 and Figure 3-7 display CMC and ROC curves for protocol II.

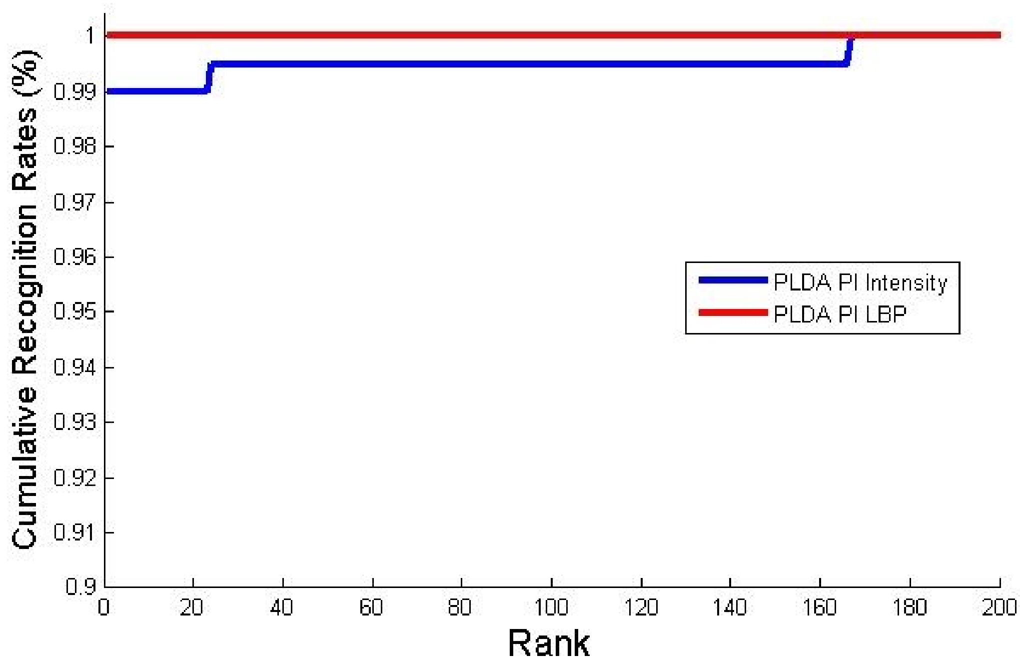

Figure 3-4: CMC curve for protocol I on HFB Face database

3.2.2 Protocol I and II of Biosecure Face Database

The Biosecure Database [50] contains 420 subjects with 12 samples taken in two sessions. Each session has 6 samples from each individual. Two samples has been captured with webcam while rest with digital camera consisting of flash and non-flash versions in each session. Biosecure database is regarded as multi-scenario and multi-environment database. I normalize all images using eye co-ordinates. All images are resized to 32x32 for experiments using intensity and LBP feature vectors. In protocol I and II, four images out of six from each session are used to create the training dataset of 300 individuals. The leftover images one from digital camera and other from webcam make up gallery and probe datasets, respectively. Table 3.2 reports the rank-1 recognition rates by comparing PLDA with other methods on both protocols.

PLDA method significantly generates very promising results on this database. It is evident from the results that this approach clearly outperforms all competing methods with very good margin. It is to be noted that none of the approaches in intermodality matching reports results on the Biosecure face database comparing

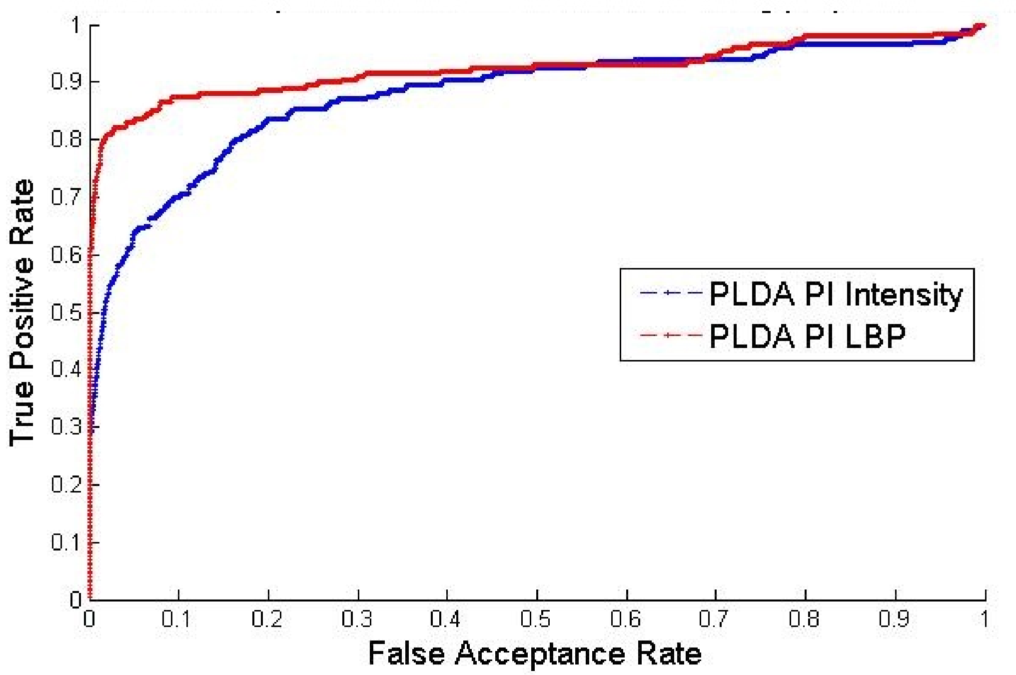

Figure 3-5: ROC curve for protocol I on HFB Face database

Table 3.2: Results for the Biosecure protocol I and II

	Recognition Rates (%)			
Methods	Intensity/PI	LBP/PI	Intensity/PII	LBP/PII
PCA [34]	20.00	22.70	25.33	22.33
LDA [33]	23.00	5.33	30.33	9.00
KDA [76]	46.67	5.33	55.33	5.33
Proposed PLDA [2]	**88.50**	**92.50**	**90.00**	**94.67**

webcam (low resolution) and digital camera (high resolution).

To further explore Biosecure database, I use all images from two sessions to make up my protocols. I select web camera images i.e. fa1 and fa2 alternatively as a query image while digital images i.e. fnf1, fnf2, fwf1 and fwf2 are used as gallery images in turn. The training set for this experiment is same as above i.e. three images from digital camera and one image from web camera have been selected. Figure 3-8 and Figure 3-9 represent CMC curves for Biosecure database where probe image is selected from web camera images and gallery image is selected from digital camera images. LBP features perform well compared to intensity / raw-pixel features as all LBP-based CMC curves in different protocol settings are skewed more to upper left

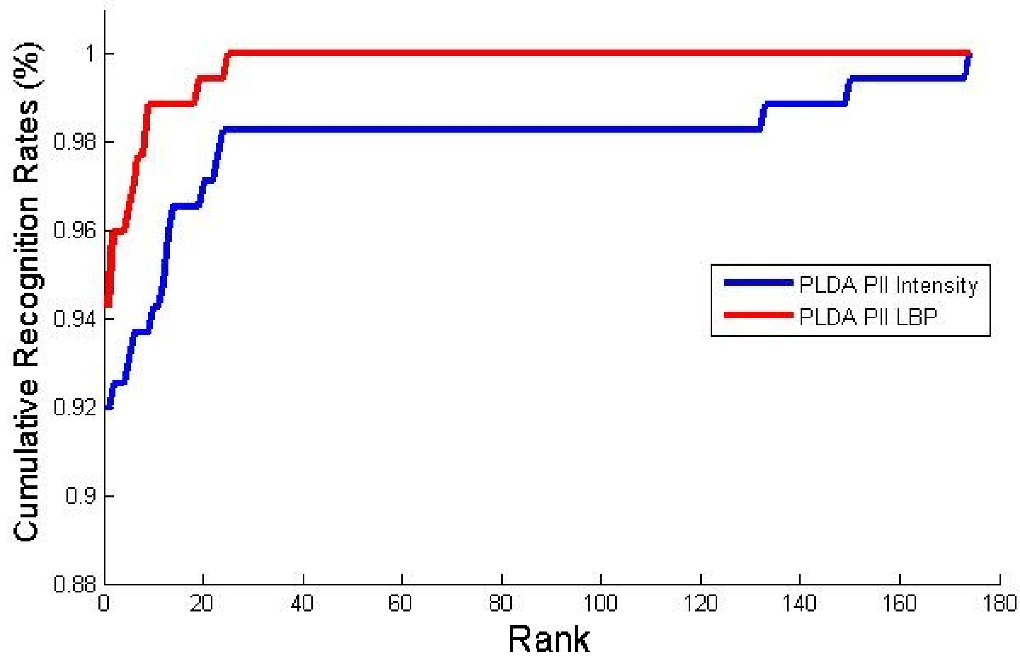

Figure 3-6: CMC curve for protocol II on HFB Face database

corner compared to intensity-based curves.

I also draw the ROC curve in fiugre 3-10 to provide another biometric performance measure to depict the comparison between LBP and intensity features using session 1 and 2 on my protocol setup. Again LBP features perform well compared to intensity features.

3.2.3 Experiment on CUHK Face Database

CUHK [51] is a publicly available dataset containing 188 subjects of two different modalities i.e face vs. sketch. In training , 88 face-sketch pairs are used while testing dataset consists of 100 pairs. All the images are cropped to 128 x 128 eye co-ordinates. The experiment using same intensity and LBP features with resizing of each image to 32 x 32 similar to HFB database experiments.

The results on this dataset is not very promising as compared to other methods but one thing is interesting to mention that all methods except LDA have considerable drop of recognition rates when LBP features are applied. But PLDA shows consistent

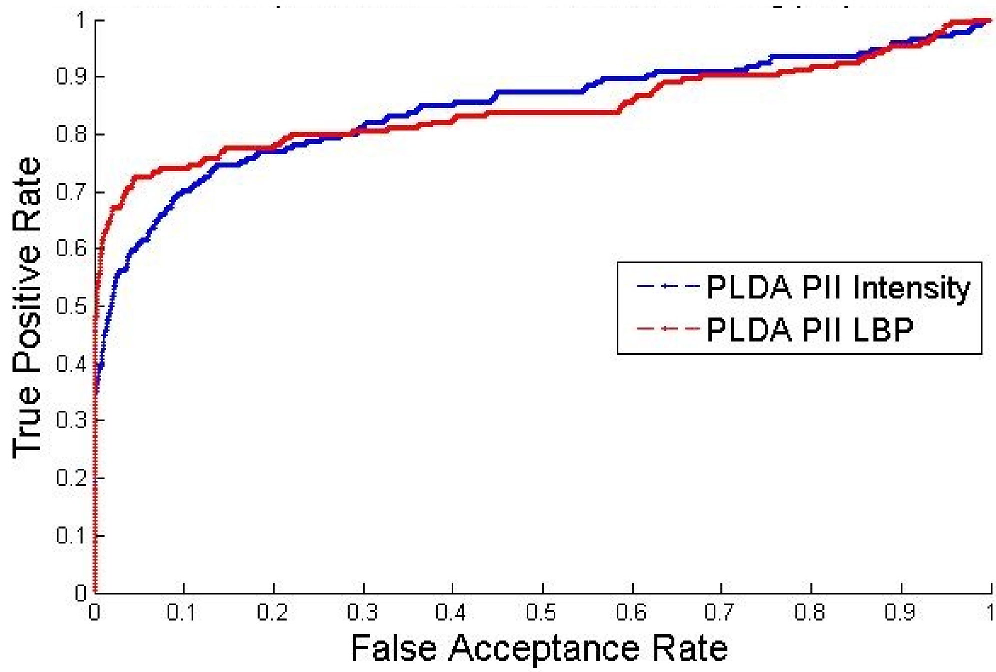

Figure 3-7: ROC curve for protocol II on HFB Face database

Table 3.3: Results for the CUHK Database

	Recognition Rates (%)	
Methods	Intensity	LBP
LDA [46]	87.00	88.00
CDFE [46]	75.00	67.00
LCSR [46]	93.00	89.00
LDSR [46]	**95.00**	**90.00**
KDSR [46]	**95.00**	**90.00**
Proposed PLDA [2]	69.00	74.00

results on LBP features as in my other databases , it has around 7% rise in recognition rates. Further, my rates are not on higher side due to less number of training samples used. It is worth while to mention that generative models require substantial training samples to improve on generalization.

3.3 Summary

In this chapter, I utilize the efficacy of PLDA method based on posterior probabilities to LDA in the intermodality face matching problem. PLDA shows some consistent

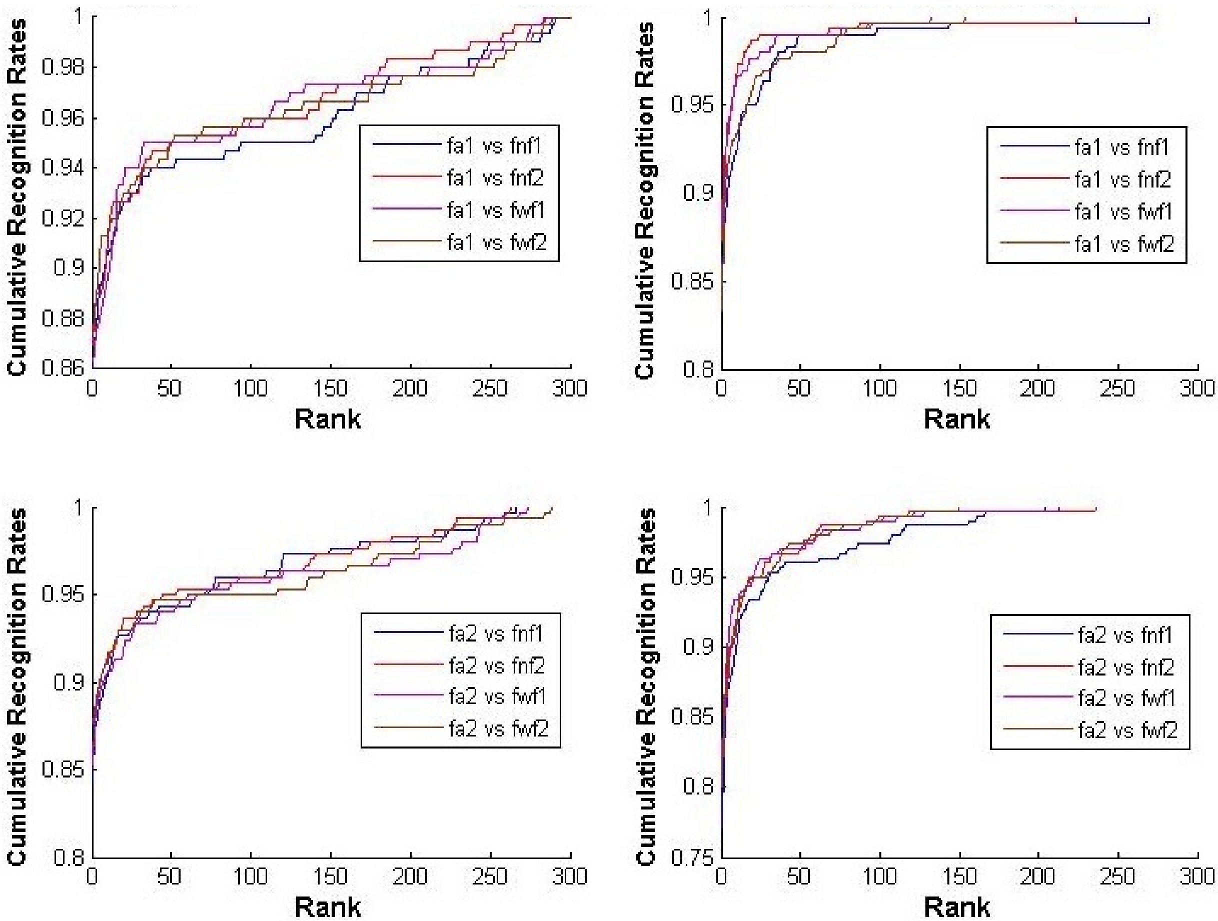

Figure 3-8: CMC curve for protocols using intensity and LBP features on Biosecure Face database Session 1

results on two different types of intermodality problems involving visual vs NIR , digital camera vs web camera and photo vs sketch. It is also to be noted that many heterogenous face matching approaches use variety of statistical learning approaches but PLDA, a generative probabilistic approach is applied first time in this domain.

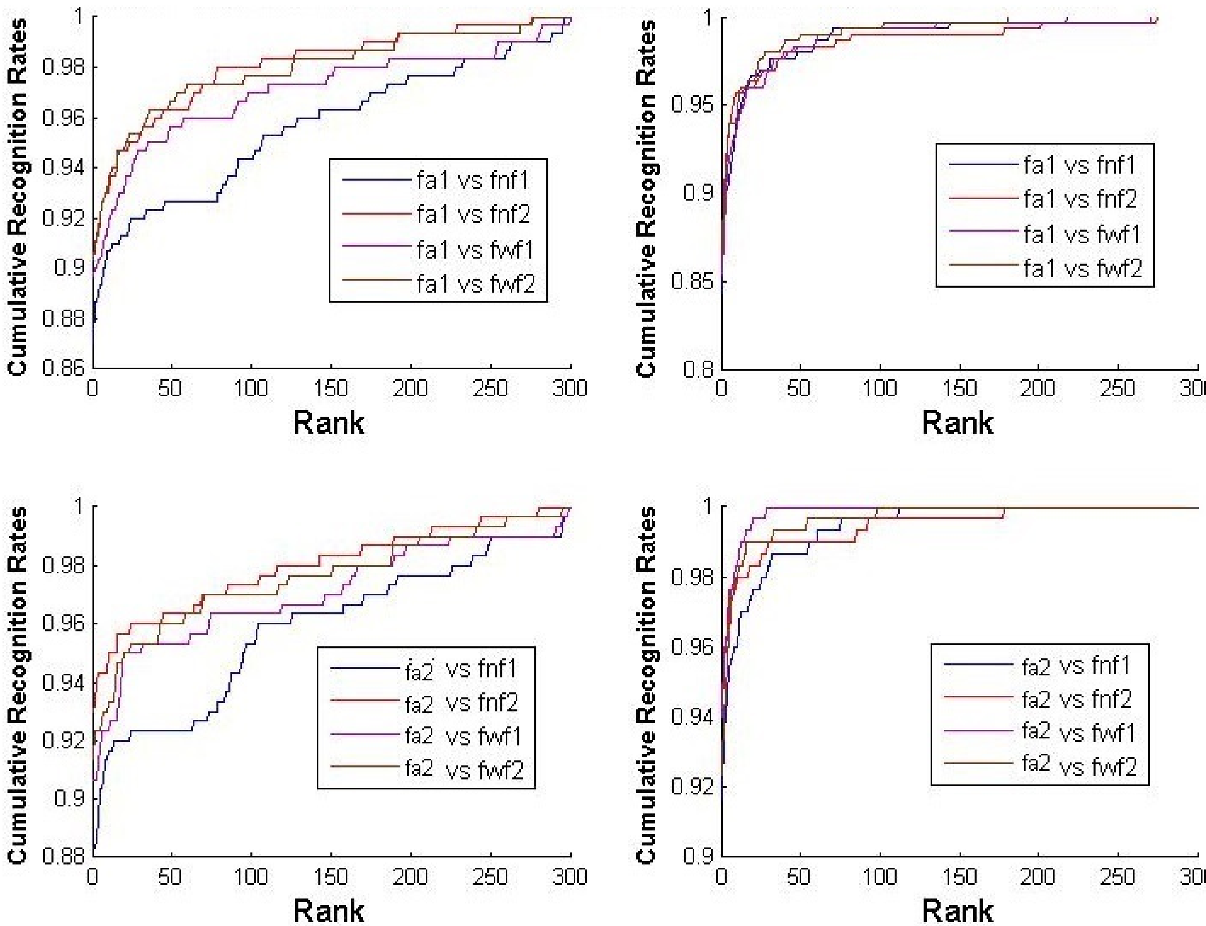

Figure 3-9: CMC curve for protocols using intensity and LBP features on Biosecure Face database Session 2

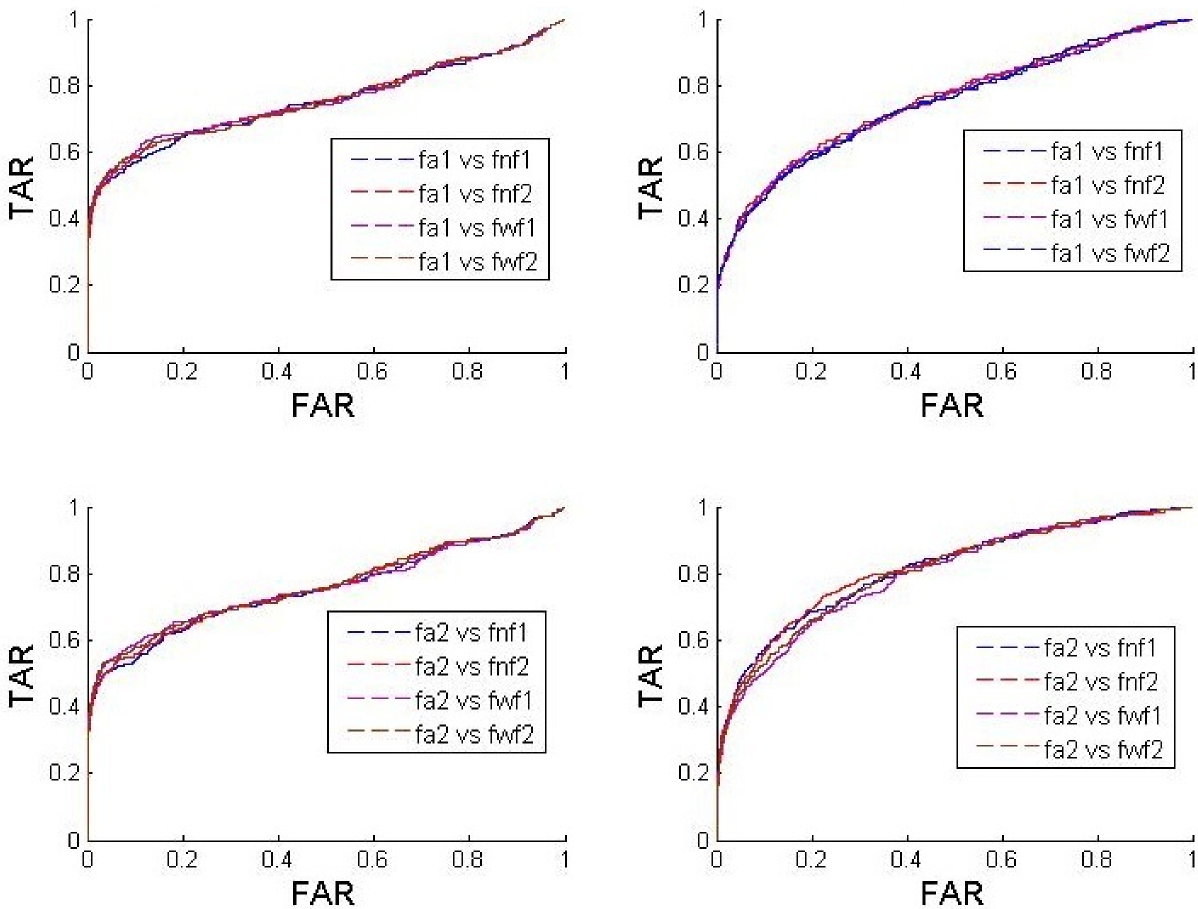

Figure 3-10: ROC curve for protocols using intensity and LBP features on Biosecure Face database Session 2

Chapter 4

Heterogeneous Face Recognition : Tied Factor Analysis using Bagging

In Heterogeneous face recognition, very few methods have been designed to solve this problem using intensity features and considered small sample size issue. In this chapter, I consider the worst case scenario when there exists a single instance of an individual image in a gallery with normal modality i.e. visual while the probe is captured with alternate modality, e.g. Near Infrared.

HFR brings three major problems namely high intra-class variability , feature gap and appearance difference. To solve these problems in inter-modality matching, biometric researchers handle it by synthesizing one modality samples into another modality to reduce the appearance difference. Y. Ma et. al. [77] synthesizes the visual image from its respective NIR by learning sparse coefficients in analysis-by-synthesis framework. Z. Zhang et. al. [78] utilizes the Lambertian reflectance model to learn the quotient image (ratio of VIS-NIR albedo). They extract NIR and quotient patches termed as multi-factors from different spectral channels and combine them with kernel approach to reconstruct the visual image.

Other efforts in this domain are to learn and design discriminative features to suppress the effects of feature gaps arisen from VIS-NIR modalities. Z. Lei et. al. [26] has proposed a new discriminant face descriptor (DFD) inspired from local binary pattern (LBP). Rather using a binary code of LBP, they extract a pixel difference matrix (PDM) by applying optimal neighborhood sampling strategy and project this PDM to form discriminant pattern vector. Finally , the dominant pattern for an image is obtained by applying clustering technique on the sets of pattern vectors. Logarithm gradient histogram feature descriptor [79] is developed based again on Lambertian model to reduce the illumination effect by simultaneously considering log-gradient, log-magnitude and spectral wavelength of lighting.

Researchers in this area have also explored variety of subspace learning approaches to tackle the appearance and feature level differences to VIS-NIR face matching. The first ever approach in VIS-NIR scenario was proposed by Lin and Tang [75]. They design common discriminant feature representation by learning projections for reference and query faces. Recently, there has been flurry of research on this active domain and variety of subspace methods have been developed. X. Huang et. al. [46] applied regularized coupled spectral regression to learn the low dimensional projections for each modality via graph embedding. Brendan and Anil [48] have proposed a new generic approach termed as prototype random subspace to handle alternate modalities presence in probe images by matching their non-linear similarities using random subspace linear discriminant analysis. The proposed method in this chapter also lies in the same domain but objective is to learn new latent subspace in which the faces from different modalities are represented by a unique latent identity variable.

In this chapter, a new technique based on Tied Factor Analysis (TFA) [32] is proposed by learning latent identity subspace from the observed samples of visual and NIR modalities as both modalities samples can be represented by a unique variable in new subspace. The proposed method identifies an existence of latent identity variable (LIV) in the new latent subspace which describes how an underlying modality invariant representation created the modality varying (observed) data. Since it is a Bayesian generative approach therefore it brings quite interesting characteristics such as careful modelling of noise, ignoring variables of least interest by marginalizing over them and providing a coherent way of comparing models using Bayesian model comparison. Original TFA method is designed to handle the small sample size problem in pose problem of homogeneous face recognition considering only one modality. I extend this approach to heterogeneous face recognition research problem but results can be misleading by only reporting on small subset of training and testing samples. To resolve this problem, I use cross-validation leave-one-out strategy in HFR scenario to remove the prejudice associated with TFA. I term this new approach as Bagging based TFA.

Following are the main contributions of this chapter

- The original TFA method works only with one training pair of same modality i.e. two images from different poses. In the proposed method, I extend TFA to handle heterogeneous modalities i.e. two images from different sensors e.g. VIS-NIR and Webcam-Digicam. Still the results from TFA can be misleading and biased as one can report results on the subset of images which produced high accuracies. In order to overcome this deficiency of traditional TFA, Bagging based TFA method is proposed to exhaustively test face databases in cross validation environment with leave one out strategy to report fair and comparable results.

The remainder of the chapter is organized as follows. In Section 4.1 , I describe tied factor analysis, its training and recognition stages. Section 4.2 discusses the proposed method bagging based TFA. Experimental setup and results are presented in Section 4.3. Section 4.4 concludes this chapter.

4.1 Tied Factor Analysis (TFA)

TFA [32] is a very interesting statistical generative model that reconstructs the observed data with smaller set of LIV variables in the presence of small sample size. It provides one to many mapping from latent identity subspace to observed space. In latent identity space, the image representation does not change with the modality and remain tied. TFA models the observed feature vector as being generated by the modality-contingent linear transformation of identity variable in the presence of Gaussian noise. In this model, the factors (linear transformation) depends on specific modality but the factor loadings are tied (constant) with the individual i.e unique latent variable represents two different modality samples (e.g. VIS-NIR) of same person.

The key property of this model is that there exits a multidimensional variable in a new subspace which represents the identity of an individual regardless of the modality. Further, if two LIVs take the same values then it corresponds to an identity of an individual and vice versa. The model never measures the LIVs directly but observed images have been generated from latent variable with its associated noise.

4.1.1 TFA Model Description

TFA [32] considers that there exists J examples of K modalities for each of I different individuals. It indicates the j^{th} image of an i^{th} individual in the k^{th} modality by x_{ijk}. TFA model is of the form

$$\mathrm{x}_{ijk} = \mathrm{F}_k \mathrm{h}_i + \mathrm{m}_k + \epsilon_{ijk} \tag{4.1}$$

where m_k represents the overall mean of the training dataset, F_k denotes the deterministic transformation function between identity and observed space, h_i represents latent identity variable associated with i^{th} individual and ϵ_{ijk} is the noise term defined as multivariate Gaussian with diagonal covariance Σ and zero mean. There is one F and m for each modality k.

The TFA broadly models actual image generation process where LIV describes the shape and structure of face, transformation function represents the camera projection process and the noise term describes the noise associated with camera during image capture plus remaining associated variables. This generative model is closely related to the factor analysis where factors, F_k, depend on the modality but the factor loadings, h_i, remain constant at each modality i.e. tied.

Formally, the TFA model can be described using conditional probabilities as

$$P_r(\mathrm{x}_{ijk}|\mathrm{h}_i) = g_x[\mathrm{F}_k \mathrm{h}_i + \mathrm{m}_k, \Sigma_k] \tag{4.2}$$

$$P_r(\mathrm{h}_i) = g_h[0, I] \tag{4.3}$$

where $g_a[b, C]$ describes a Gaussian in a with mean b and covariance C. Equations 4.3 is defining simple priors on h_i.

In TFA model, the only known parameter is the observed images while the rest θ $= \mathrm{F}_{1...K}, \mathrm{m}_{1...K}, \Sigma_{1...K}$ are all unknown. If I know h, then learning parameters F , m and Σ will be quite easier. But, unfortunately, all the right hand side parameters of proposed TFA model in Equation 4.2 is unknown.

Fortunately, for this chicken-egg problem, one can take advantage of Expectation and Maximization Algorithm. It iteratively maximizes the likelihood of parameters alternately in each iteration. The E step finds the unknown identity variables h by calculating posterior probabilities over fixed parameter values i.e calculating expected values of h_i for each individual i by using data for individual across all modalities $\mathrm{x}_{i..}$.

The first two moments of Expectation (E) steps are

$$\begin{aligned} E[h_i|x_{i..}] =& \Big(I + \sum_{j=1}^{J}\sum_{k=1}^{K} F_k^T \Sigma_k^{-1} F_k\Big)^{-1}. \\ & \sum_{j=1}^{J}\sum_{k=1}^{K} F_k^T \Sigma_k^{-1}(x_{ijk} - m_k\Big) \end{aligned} \tag{4.4}$$

$$\begin{aligned} E[h_i h_i^T|x_{i..}] =& \Big(I + \sum_{j=1}^{J}\sum_{k=1}^{K} F_k^T \Sigma_k^{-1} F_k\Big)^{-1} \\ & + E[h_i|x_{i..}]E[h_i|x_{i..}]^T \end{aligned} \tag{4.5}$$

In M step, the algorithm maximizes the lower bound on the parameters $\theta = F, m, \Sigma$ for each modality k. The update rules for this step are

$$\begin{aligned} \tilde{\mathrm{F}}_k =& \left(\sum_{i=1}^{I}\sum_{j=1}^{J} \mathrm{x}_{ijk}\mathrm{E}[\tilde{\mathrm{h}}_i|\mathrm{x}_{i..}]^T\right). \\ & \left(\sum_{i=1}^{I}\sum_{j=1}^{J} \mathrm{E}[\tilde{\mathrm{h}}_i\tilde{\mathrm{h}}_i^T|\mathrm{x}_{i..}]\right)^{-1} \end{aligned} \tag{4.6}$$

$$\begin{aligned} \Sigma_k =& \frac{1}{IJ}\sum_{i=1}^{I}\sum_{j=1}^{J} diag \\ & \Big[\mathrm{x}_{ijk}\mathrm{x}_{ijk}^T - \tilde{\mathrm{F}}_k\mathrm{E}[\tilde{\mathrm{h}}_i|\mathrm{x}_{i..}]\mathrm{x}_{ijk}^T\Big] \end{aligned} \tag{4.7}$$

where diag corresponds to retain the diagonal elements of the matrix.

4.1.2 TFA Learning Process

After learning TFA model parameters, it is good practice to verify and confirm that model has learned relationship between alternate modalities. I use HFB face database containing 200 individuals having samples of VIS-NIR modalities. I select 133 individuals for training and 66 for testing at each modality. Each image is resized to 70x70 and its raw pixel values are concatenated to make long observation vector. I learn parameters $\theta = F, m, \Sigma$ for each modality k i.e. VIS and NIR by applying 10 iterations of EM algorithm. The only tunable parameter in the experiment is F. To visualize the learning stage, I first represent the original and reconstructed images in Figure 4-1 to verify the accurate learning of VIS-NIR parameters. The original and its reconstructed version of both modalities resemble closely to each other.

Figure 4-1: 1st and 3rd columns contain original VIS-NIR images and 2nd and 4th columns contain reconstructed images of HFB

To further check learning of TFA, I predict the NIR images from its respective VIS images by acquiring latent identity variable h from its posterior distribution. Figure 4-2 confirms the claim that both modalities share the same LIV as reconstructed NIR is closely matching its counterpart.

Figure 4-2: 1st and 2nd columns contain original VIS-NIR images and 3rd column contain reconstructed NIR image using tied factor model with 64 factors

4.1.3 Recognition Stage

After learning model parameters, the approach in recognition stage is to match gallery and probe faces where it can be represented by exactly the same values of the identity variable. In TFA formulation due to noisy observations, I integrate out LIV's to get final decision which doest not depend on single point estimate of h. The recognition stage of TFA compares the likelihood of the data under N different models which is denoted by $M_{1...N}$. In a closed set identification, the n^{th} model M_n represents the case where probe face x_p matches the n_{th} gallery face so n^{th} identity variable h_n is responsible of generating probe feature vector i.e. $h_p = h_n$.

The evidence for the model M_n i.e. match can be given as

$$
\begin{aligned}
&P_r(\mathrm{x}_{1,\ldots,N}, \mathrm{x}_p|\mathrm{M}_n) \\
&\quad = \int P_r(\mathrm{x}_{1,\ldots,N}, \mathrm{x}_p, \mathrm{h}_{1,\ldots,N}, \mathrm{h}_p|\mathrm{h}_p = \mathrm{h}_n) d\mathrm{h}_{1,\ldots,N} \\
&\quad = \int P_r(\mathrm{x}_1|\mathrm{h}_1) P_r(\mathrm{h}_1) d\mathrm{h}_1 \ldots \\
&\quad \int P_r(\mathrm{x}_n, \mathrm{x}_p|\mathrm{h}_n) P_r(\mathrm{h}_n) d\mathrm{h}_n \ldots \\
&\quad \int P_r(\mathrm{x}_N|\mathrm{h}_N) P_r(\mathrm{h}_N) d\mathrm{h}_N
\end{aligned} \tag{4.8}
$$

The evaluation of above integrals is basically the reformulation of generative equation as a standard factor analyzer and results in a composite system:

$$
\begin{bmatrix} \mathrm{x}_1 \\ \mathrm{x}_2 \\ \vdots \\ \mathrm{x}_Q \end{bmatrix} = \begin{bmatrix} \mathrm{F}_1 \\ \mathrm{F}_2 \\ \vdots \\ \mathrm{F}_Q \end{bmatrix} \mathrm{h}_i + \begin{bmatrix} \mathrm{m}_1 \\ \mathrm{m}_2 \\ \vdots \\ \mathrm{m}_Q \end{bmatrix} + \begin{bmatrix} \epsilon_1 \\ \epsilon_2 \\ \vdots \\ \epsilon_Q \end{bmatrix} \tag{4.9}
$$

The above formulation can be rewritten as

$$
\mathrm{x}' = \mathrm{F}'\mathrm{h}_i + \mu' + \epsilon' \tag{4.10}
$$

This now transforms into standard factor analyzer whose likelihood is well established i.e $g_{x'}[m', F'F'^T + \Sigma']$

The posterior probabilities of probe matching galleries after calculating evidence for each model can be given as

$$
\begin{aligned}
&P_r(M_n|\mathrm{x}_{1,\ldots,N}, \mathrm{x}_p, \theta) \\
&\quad = \frac{P_r(\mathrm{x}_{1,\ldots,N}, \mathrm{x}_p|M_n, \theta) P_r(M_n)}{\sum_{m=1}^{N} P_r(\mathrm{x}_{1,\ldots,N}, \mathrm{x}_p|M_m, \theta) P_r(M_m)}
\end{aligned} \tag{4.11}
$$

4.2 Proposed Bagging based Tied Factor Analysis for heterogeneous face recognition

In this section, I discuss the proposed bagging TFA method. I first put some light on Bagging as this is the main concept used in TFA.

4.2.1 Bagging

Bagging [80], the popular machine learning method, generates many training subsets. Each training subset is selected by randomly generating subsets where each sample is selected with replacement and equal probability. A prediction method is applied to each training subset to get base model. A new test sample is processed by all models to classify that new test sample. The final decision of the class is normally obtained by majority voting but other combining rules such as mean, product and average are also used in bagging. The main advantage of bagging is that aggregation often performs better than base models and results in reduction of variance.

4.2.2 Bagging based TFA

TFA is designed to solve small sample size (SSS) issue in pose related problems of homogeneous face recognition. In this proposed method, I extend TFA to heterogeneous recognition case to handle SSS in the presence of alternate modalities. But, it can only report correct matches for only two pairs of heterogeneous modalities. One VIS-NIR pair consisting of N subjects is used for training while other VIS-NIR pair of M subjects are used for testing. So, the results from TFA classifier may be misleading and biased as one can produce results on those smaller subset of database that produces high recognition rates. The extensive evaluation of heterogeneous face databases is therefore taken into consideration to report fair and comparative results with some of the state-of-art-methods [46].

Due to this limited nature, I have come up with an idea to extend its efficacy by training TFA with all training samples applying leave one out strategy and testing it with left out sample with all learned bagged TFA ensembles. In the setup, I have used lth pairs of specific VIS-NIR images $\mathrm{X}_v^l, \mathrm{X}_r^l$ for N training subjects to learn the TFA model parameters for each modality i.e $\mathrm{F}_v^l, \mathrm{m}_v^l, \Sigma_v^l$ and $\mathrm{F}_r^l, \mathrm{m}_r^l, \Sigma_r^l$. By applying cross validation scheme using leave one out strategy, I test the TFA with q pairs of M test subjects left out in training with learned TFA model parameters. Lastly, I combine the scores from different base models by applying sum rule to get the final decision.

4.3 Experiments and Results

The effectiveness of TFA is tested on two intermodality scenario i.e Visual vs. NIR and low-resolution vs. high resolution. For testing , intensity or raw features are employed and rank one recognition rates and ROC curves are presented.

4.3.1 HFB VIS-NIR Face Database

To exhaustively test HFB using Bagging TFA, I employ three different protocols out of them protocol I (PI) and protocol II (PII) adapted from coupled spectral regression variants [46] , [81] and [82]. Following setting in [83], the third protocol (PIII) contains non-overlapping 100 train and test subjects selected randomly. The samples of VIS-NIR database are of the size 128 x 128 cropped using eye co-ordinates. For intensity features, all images are resized to 32 x 32. All the results are reported on 64 factors bagging based TFA. Table 4.1 reports rank-one recognition rates comparison of proposed method with other state-of-art methods.

Table 4.1: Results on HFB Evaluation protocol I , II and III

	Recognition Rates (%)			
Method	Intensity/PI	Intensity/PII	Method	Intensity/PIII
PCA+CCA [46, 83]	95.42	51.09	PCA+CCA [46, 83]	26.70
LCSR [82]	97.48	75.65	CSR [82, 83]	38.92
KCSR [82]	97.34	73.06	TCA [83]	0.21
ICSR [81]	98.54	77.53	tPCA [83]	3.16
LDSR [46]	97.54	73.96	SDA [84]	38.30
KDSR [46]	98.34	77.04	THFM [83]	20.04
Proposed Bagging TFA	**99.35**	**90.72**	Proposed Bagging TFA	**45.45**

It has been observed that proposed bagging based TFA reports quite consistent results on all protocols using intensity based features. In protocol I , it has marginal improvement compared to other methods. In protocol II, it achieves the gain of 17-25% on all reported methods. Protocols I and II have overlapping subjects which may not reflect the real world scenario. Therefore, protocol III has been created to effectively test HFB database in the presence of non-overlapping subjects during training and testing. It is worth to mention that results reported on protocol III by all methods use local patch-based learning of intensity features while the proposed method extract intensity features from all images as holistic-based learning. There is definitely a good margin of improvement if I apply patch-based approach to the proposed bagging based TFA. Figure 4-5 shows the receiver operating characteristics (ROC) curve plotting False Acceptance Rate(FAR) vs. Genuine Acceptance Rate (GAR) for different configurations on HFB face database and it is evident that bagging based TFA, the colored plot, performs better than all individual TFA based models in all protocols.

4.3.2 Biosecure Face Database

The Biosecure Database [50] , multi-scenario and multi-environment database, has 420 subjects with total 12 samples of each individual taken in two sessions. Two samples has been obtained with web camera and the rest with digital camera having two samples each of flash and non-flash versions in each session. All images for the experiment are normalized using eye co-ordinates and resized to 32x32 for intensity feature vectors. For protocol I , webcam and digitalcam images are used alternately in training and testing subjects. In this case, the subjects in test set are partially overlapped with training set. Similar to previous protocol setup, protocol II is also created with no overlap between train and test dataset. For protocol II, 200 are used for training and 100 for testing. The recognition rates achieved for proposed method are 94% and 55% for protocol I and II respectively. The superiority of the proposed method over base TFA models is presented in Figure 4-6 which reflects the ROC curve for protocol II showing FAR vs. GAR plot on different thresholds.

4.4 Summary

Tied Factor Analysis, a generative Bayesian approach in homogeneous face recognition, has produced outstanding results in the presence of extreme pose variation when there exist only one training sample for each individual. In this chapter, I extend Tied Factor method utilization in more challenging case of heterogeneous face matching when there exists only one training sample of different modalities. But, TFA's result can be considered unfair in the presence of small sample size as any one can report result on small subset of data set which may produce high recognition rates. Therefore, extensive evaluation on HFB and Biosecure database are carried out with leave-one-out cross validation using Bagging based TFA. It is evident that the proposed method not only improves the results due to bagging compared to base TFA models but also effectively outnumber other state-of-the-art approaches with good margin by just using intensity features.

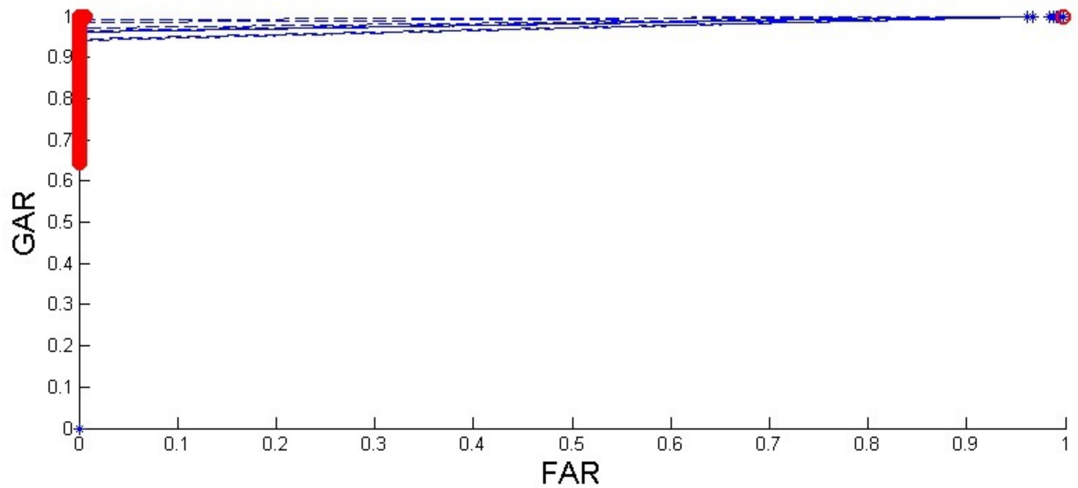

Figure 4-3: ROC Curves for Protocol I HFB Database

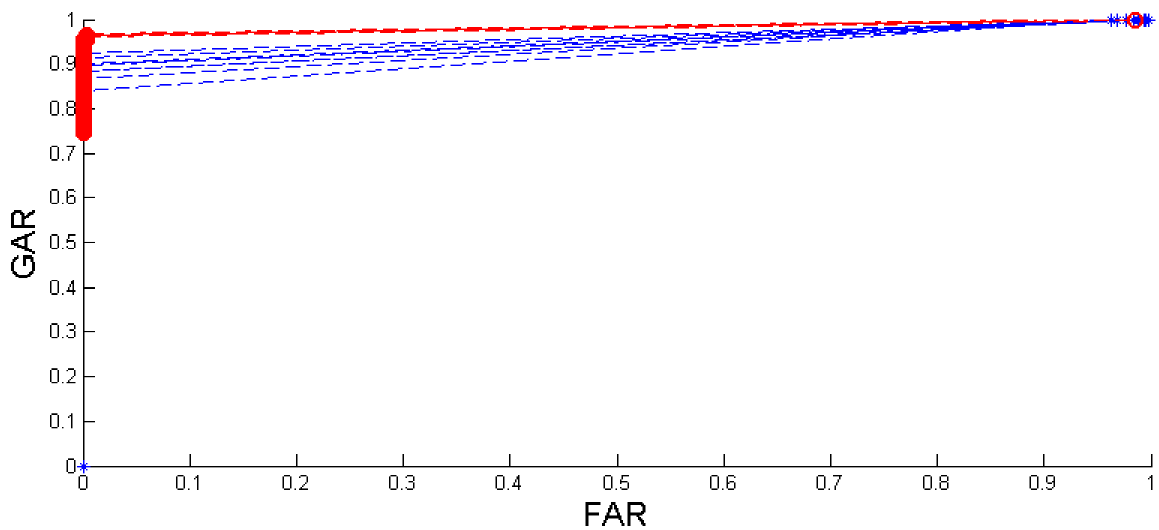

Figure 4-4: ROC Curves for Protocol II HFB Database

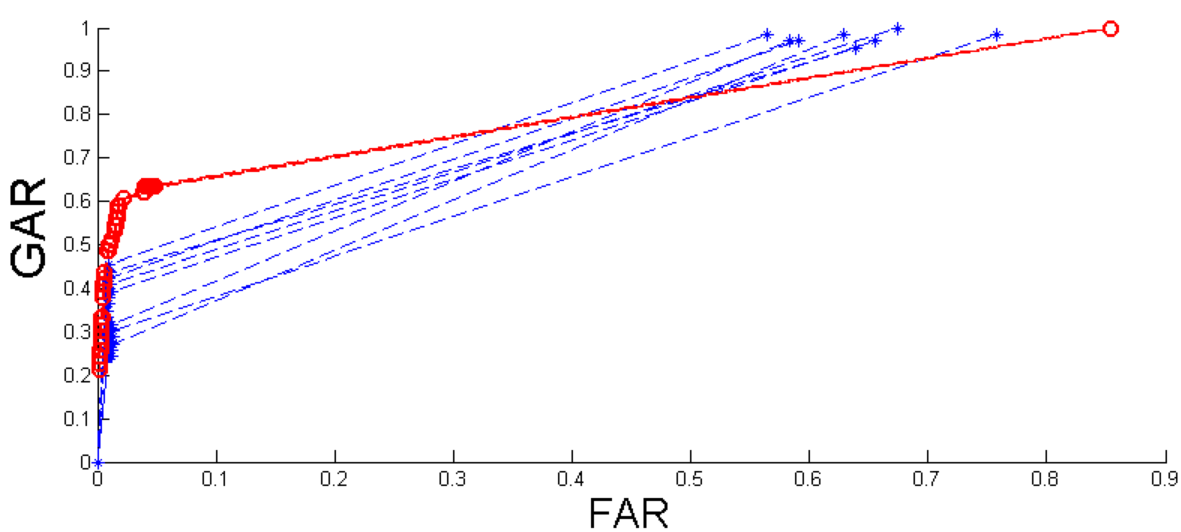

Figure 4-5: ROC Curves for Protocol III HFB Database

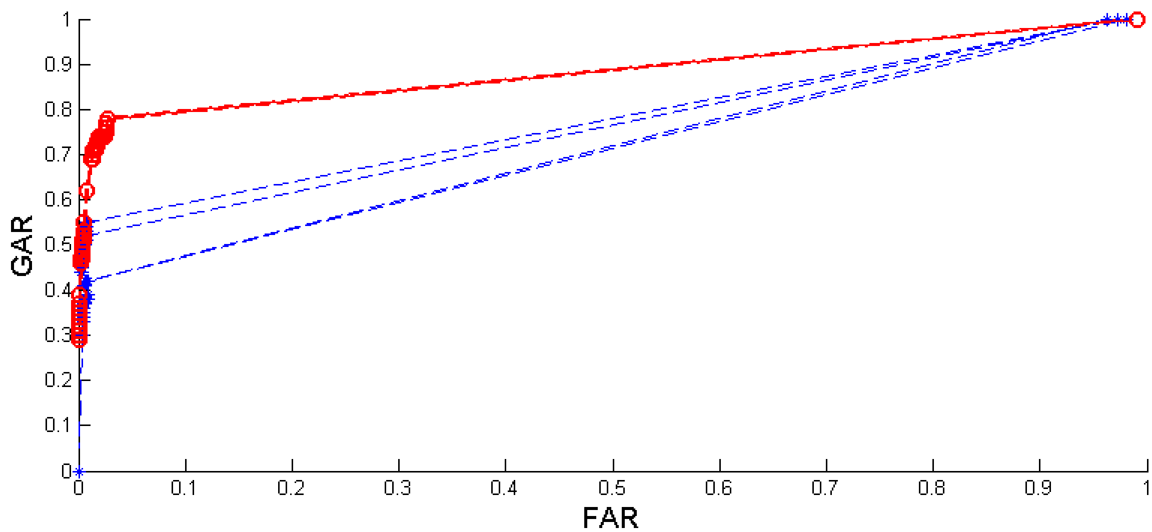

Figure 4-6: ROC Curves for Protocol II Biosecure Database

Chapter 5

Modality Identification for Heterogeneous Face Recognition

5.1 Introduction

The flurry of research to solve heterogenous face recognition (HFR) problem has brought some interesting state-of-the-art methods in this challenging research problem. Visual, Near Infra Red, digital-cam, web-cam, sketch, thermal etc [48] are types of heterogeneous modalities of HFR domain. The majority of older and recent approaches assume that the sensor/modality identity is known prior to its application at the recognition stage. Although some of the methods are claimed to be automated HFR, in reality these methods clearly use the notion of known modality identity prior to the recognition process [48, 47, 85]. This leads to a major drawback for the automation of HFR systems as real world scenarios cannot be reflected. No effort has been made to develop a fully automated inter-modality face matching approach. The reason is that there does not exist such a mechanism or a module which can identify sensor/modality types.

Until now, the research community working on HFR problem has managed to maintain different types of HFR face databases individually [49, 50, 51]. Researchers have applied their proposed methods on each of the database separately and reported the performance accordingly. Some methods in HFR [75, 46, 48] claim to be robust

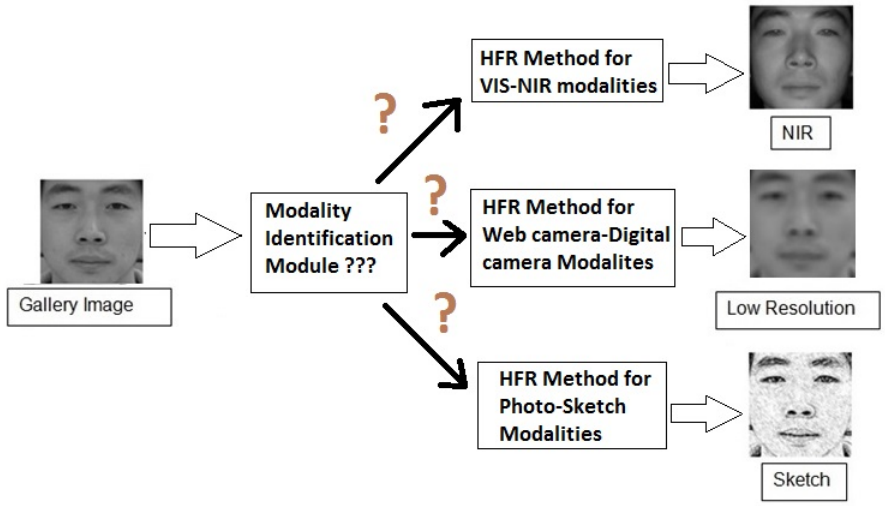

Figure 5-1: The matching of probe with template face in Gallery. Probe face modality can be NIR, low resolution, sketch etc. while Gallery image has high resolution. Real world scenario where the importance of modality identification module is monumental.

by reporting good performance rates in more than one modality e.g VIS versus NIR, photo versus sketch. But with the recent advancement of face recognition technology, a need will arise to identify the modality of face for bringing automation to this field. A FR system should be smart enough to recognize the modality of the face and forward the request of matching probe face with template in gallery to the correct HFR method handling the identified modality type of the face.

Figure 5-1 represents real world scenario where the image of the person in the gallery needs to be matched with correct probe modality for example Visual(VIS) versus Near Intra Red (NIR), VIS versus Sketch, VIS versus Low-Resolution, VIS versus Thermal. But currently, the module to identify modality type is missing in the face recognition pipeline as all HFR methods assumes the modality type is known prior to the recognition. In this figure, I use an image pair from HFB database [49] maintained by CASIA dataset (deals with visual and NIR modalities) and generate low-resolution and sketch image from visual to provide the visualisation of real world set-up. In reality, there exists no heterogeneous face database which deals with majority of modality types. The current direction and development of HFR methods demands a need of common and universal inter-modality database which shall contain all type of modalities for rigorous testing of the proposed state-of-the-art heterogeneous face recognition methods.

5.1.1 Motivation for Modality Identification

Inspired from camera identification and liveness identification, heterogeneous face recognition also requires similar attention towards modality identification. In the future, FR either homogeneous or heterogeneous will move towards automation and will result in deployment of smart gadgets. The automation of FR technology requires no manual intervention or manual identification of source modalities. Further, due to the complexity of face manifold, there is a very small chance of having a single robust method to solve the majority of problems in face recognition research. Therefore, in the future, I may expect devices using methods specialized in solving one particular problem of HFR. In fact identifying face modalities can be as important as recognizing different poses, camera source identification and face liveness detection in the biometric systems. In this chapter, a novel method is proposed for modality identification which is inspired from sensor pattern noise (SPN) estimation based approaches in camera / source identification [86]. The proposed method is simple and computationally effective to implement. The following are the main contributions of this paper

- The modality identification problem is basically ignored in the majority of state-of-the-art intermodality face recognition methods and all methods have assumed the availability of the image modality information. To the best of my knowledge, this is the first study that aims to consider the modality identification problem in Heterogenous Face Recognition scenario.

- The identification of modalities has similar importance in biometric recognition systems when compared with pose identification, camera identification, liveliness detection, gesture recognition etc. The module which can identify the modality types is missing in face recognition pipeline. This paper puts emphasis on this important module which will result in bringing the fully automated

HFR methods.

The remainder of the paper is organized as follows. Section 5.2 provides information on HFR methods dealing with more than one modality. Section 5.3 discusses the motivation of the proposed modality identification by considering the concept of camera identification. Section 5.4 discusses face anti-spoofing measures that will be used to provide fair comparison between proposed approach and face anti-spoofing approach. Experiment set-up and results are presented in Section 5.5 followed by discussion in Section 5.6. Finally, conclusion is provided in Section 5.7.

5.2 HFR Methods Dealing More Than One Modality

The first method of HFR to investigate performance measures on two different modalities i.e. photo vs sketch and visual vs. NIR was proposed by D. Lin et. al. [75]. The authors proposed Common Discriminant Feature Extraction which transforms different modalities simultaneously to a common feature subspace. Their algorithm formulates the learning objective by incorporating both the empirical discriminative power and the local smoothness of the feature transformation. The complexity of the model is controlled through the smoothness constraint, thus reducing the risk of overfitting and subsequently enhancing the generalization capability.

Coupled discriminant analysis proposed by Z. Lei et. al. [47] is another example of robust method which successfully handles more than one modality. The effectiveness of this method is shown by carrying extensive experiments on three cases of HFR i.e. high vs. low image resolution, digital photo vs. video image, and visible light vs. near infrared. This method used all samples from different modalities to represent the coupled projection in order to extract discriminative information. It performs quite well in majority of HFR scenario but does not report results on photo vs sketch.

X. Cai et al. [87] proposed a method based on coupled latent least squares regression. The method assumes that images from different modalities are generated from latent ideal object and the modality invariant information is formulated using least square regression. Extensive experiments on visible light vs. near infrared, and photo vs. sketch validates the efficacy of this method.

B. F. Klare et al. [48] proposed a framework in which both the probe and gallery images are represented in terms of nonlinear similarities to a collection of prototype face images. This non-linear representation of the prototype is improved using a projection of compact features into linear discriminant subspace. Finally, the authors use random subspace sampling to reduce the effect of small sample size. This approach reports recognition rates on four different modalities NIR vs. Visual, Thermal vs. Visual, Viewed-Sketch vs. Visual, Forensic-Sketch vs. Visual. Although, this method appears to handle many different scenarios of inter-modality face recognition but it also relies on the prior knowledge of each type of modality.

A method named discriminative spectral regression (DSR) maps heterogeneous face images into a common discriminative subspace in which robust classification can be achieved. The method proposed by X. Huang [46] transformed the subspace learning into a least squares problem. The method further introduce two regularization terms that validate this proposed method. This method uses two different HFR scenarios i.e. VIS vs NIR and Photo vs sketch.

In the past few years, a few methods have been proposed [88, 85, 89] which report outstanding results on Visual vs. NIR face databases e.g. HFB etc. This supports my claim that with more research on HFR problem, new and effective methods in this domain to handle specifically one type of modality as these methods report almost 100% accuracy will be generated. The idea to handle HFR modalities with methods specialized in particular inter-modality case would increase the demand of module to identify the type of modality so that request of recognizing probe will pass to correct recognition method. Figure 5-2 represents the HFR pipeline [54] where the modality identification module is missing in all state-of-art methods in inter-modality face matching problem.

Very recently in [90], the authors have proposed a method to model the non-linear relationship of heterogeneous faces by using Restricted Boltzmann Machines

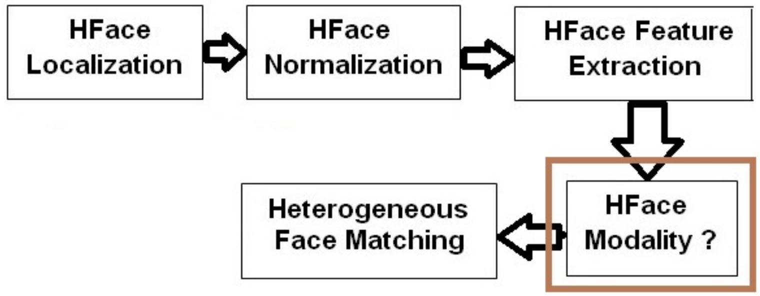

Figure 5-2: Block Diagram reflecting the missing module of modality identification in Heterogeneous face recognition pipeline (HFace : Heterogeneous Face).

(RBMs) to learn a shared representation locally to remove the heterogeneity around each facial point. The authors reported comparable results of new variant of CASIA HFB database [49] named as CASIA v2.0 [91].

All above-mentioned state-of-art methods in HFR assume that modality of an image is known prior to their application in their methods. This currently hampers in bringing automated intermodality face recognition. To the best of my knowledge, there is currently no effort being made to solve this problem.

5.3 Image Sharpening based Modality Pattern Noise (MPN) estimation for modality identification

In this section, I will discuss my proposed framework for modality identification. Figure 5-3 shows my proposed framework. Extracting and Identifying Modality Pattern Noise (MPN) are the main components of the proposed framework and are described below.

5.3.1 Unsharp Masking pre-processing Tool

The first block in my proposed framework is image sharpening which is used as pre-processing tool before calculating specific Modality pattern noise (MPN). Image sharpening is an important step to improve the identification of source modality.

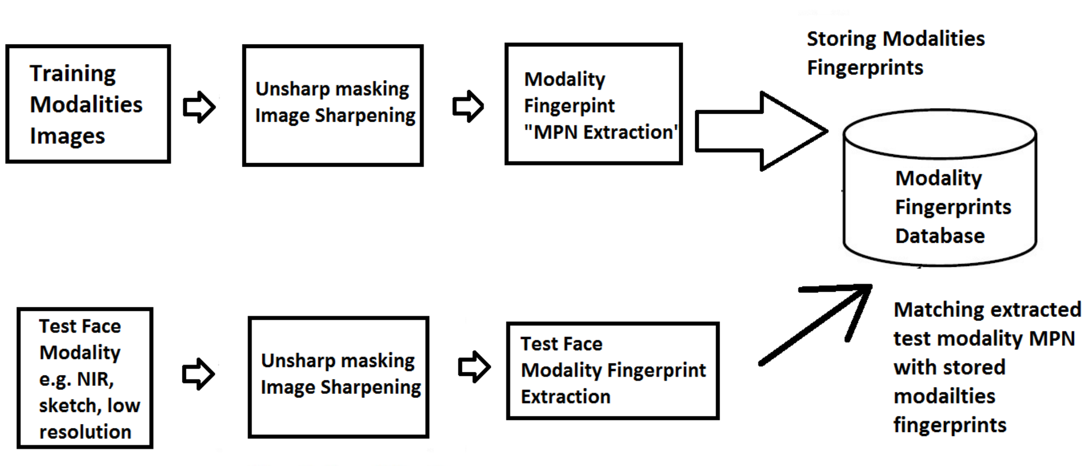

Figure 5-3: Proposed Modality Pattern Noise (MPN) framework for modality identification in Heterogeneous face recognition.

Generally, the image enhancement can be classified into two broad categories. The first one aims to modify a grey value of each pixel based on statistical information of the image. The second one is applied by actually separating the low and/or high frequency components of the image. Each signal component is manipulated separately and finally, both components are recombined with the different weights. The Unsharp Masking (UM) technique is adopted by applying the Laplacian filter [92] to enhance the source modality identification performance. UM is used to amplify the high frequency content of the MPN, hence strengthening its presence in the sample images for efficient face modality identification. In this thesis, I employ both Basic MPN and Phase MPN alongside unsharp masking to get the modality fingerprint of different sensors in training stage and compare the residue noise of probe with camera fingerprint to reach the final decision. Figure 5-4 displays images from three heterogeneous face databases original images and unsharp making images.

5.3.2 Modality Pattern Noise Estimation

The second block in my proposed approach is modality pattern noise (MPN) estimation. MPN estimation is inspired from sensor pattern noise (SPN) estimation based approaches in camera / source identification work. The importance of source

Figure 5-4: Heterogeneous face databases original images in first row and pre-processed images with unsharp masking tool in second row.

identification and forgery identification in multimedia forensics [93] has attracted lot of researchers and several novel methods are proposed to solve these problems. Source identification methods try to identify different devices e.g. camera, scanners, mobile phones etc. While forgery detection methods try to discover any evidence of tampering in images, camera identification methods used lens aberration, sensor imperfections, sensor pattern noise (SPN), color filter array interpolation, binary similarity measures, image features etc. Various defects in the manufacturing process of imaging sensors create noise in the sampled pixel values. There is a direct relationship between this noise and manufacturing defects of sensor so it can be utilized to forensically classify make and model of digital camera. In natural images, the dominant part of the pattern noise is the photo-response non-uniformity noise (PRNU) and it results due to different sensitivity of pixels to light caused by the inhomogeneity of silicon wafers and imperfections during the sensor manufacturing process termed as pixel non-uniformity (PNU). SPN based camera identification methods using PNU is being investigated recently compared to other techniques [94, 95, 96]. The reason behind is that SPN is able to identify the source of an image even if two cameras have the same brand and model. However, the accuracy of SPN estimation can be relied on several factors; firstly the higher the number of training, the better accuracy for SPN estimation. Furthermore, SPN could be weak in small size of images or regions that contain saturated background with dark regions [97].

Digital camera identification was first addressed by Lukas et al. [94] termed in my literature as Basic SPN. In order to identify digital image source, basic SPN utilized PRNU to extract camera specific SPN. The camera fingerprint or reference

SPN is computed by averaging residue of images from same camera using wavelet based de-noising filter. Several enhancements with respect to identification accuracy were proposed. In [97], a unified framework for device identification is proposed by estimating maximum likelihood from the simplified model of the sensor output. The main difference between [97] and [94] is that the former method use a smaller number of images to estimate the fingerprint. In [93], the false positive rates achieved during the decision process are reduced by incorporating digital camera's demosaicing characteristics. The camera fingerprint is greatly affected by the scene details which in turn affects the correct identification results. In [96] suggests that scene detail attenuation to camera fingerprint n can be reduced by assigning less weights to affected components in Digital Wavelet Transform.

Another method to enhance camera reference SPN in the frequency domain was proposed by Kang et al. [95] termed as Phase SPN aiming to remove the interference from scene details and camera signal processing. They proposed a correlation over circular correlation norm (CCN) as SPN extracted from digital images has been proved to be a unique fingerprint of digital cameras.

In proposed modality pattern noise estimation method, I apply Lukas et. al [94] approach which referred in the paper as Basic MPN using modality patter noise to extract face modality-specific MPN for identifying modality source e.g. VIS, NIR, photo, sketch etc. The modality fingerprint or reference MPN is computed by averaging the residues of modality face images from the same modality using wavelet based de-noising filter.

To compare my basic MPN, I have applied another method to enhance the modality reference MPN in the frequency domain proposed by Kang et al. [95] termed as Phase MPN aiming to remove the interference from scene details and camera signal processing. They proposed a correlation over circular correlation norm (CCN) as MPN extracted from face images has been proved to be a unique fingerprint of face modalities. The noise residue extracted from an image is [86]

$$W_j = I_j - F(I_j) \tag{5.1}$$

where $F(I_j)$ represents the de-noised image.

Then W_j is whitened first in the frequency domain and has constant Fourier magnitude coefficients except that its direct current Fourier coefficient equals zero. The phase only component of noise residue is calculated [86]

$$\begin{aligned} W_j &= DFT(W_j), \\ W_{\phi j} &= W_j/|W_j|, \end{aligned} \tag{5.2}$$

where $|W_j|$ is the Fourier magnitude of W_j and DFT is the Discrete Fourier Transform.

The camera fingerprint based on MPN [86] is calculated by averaging the phase component only of all images from one camera and performing inverse DFT

$$y = real\left(IDFT\left(\frac{\sum_{j=0}^{L-1} W_{\phi j}}{L}\right)\right) \tag{5.3}$$

5.4 Comparison of proposed MPN approach to Liveness Detection / Face Anti-spoofing Measures

In this section, I aim to highlight face anti-spoofing measures which will be later used in experiment section 5.5 to fairly compare myr proposed approach to some of state-of-the-art approaches in liveness detection research domain.

Biometric attacks i.e. photo attack using someone's facial image from the Internet or by some other false acquisition is a major threat to biometric systems. Imposters can use falsely acquired photos to gain access to the biometric system by presenting photos to live camera. There is a greater need to overcome this challenge by identifying the face image liveness. Liveness detection is not only limited to face biometric but other biometric traits e.g. fingerprint, iris etc. Liveness detection methods try to identify some form of human activity to prevent spoofing. To develop anti-spoofing methods in face biometrics are very limited although face recognition either homogeneous or heterogeneous systems is a active research topic for long time but still vulnerable to spoofing attack.

Despite the initial work on face recognition anti-spoofing dates almost a decade ago [98] but most recently major a contribution in this area is carried out by TABULA RASA European project [99, 100, 101]. The research of face anti-spoofing can be categorised as feature, sensor level and score level. The feature level work can be sub categorised further as static or dynamic. In dynamic feature level approach, it relies on the detection of motion over a face video sequence. The representative work in this domain are eye blinking detection [102], face motion detection [103] and dynamics of facial texture [104]. Static feature level approaches are based on the analysis of the face texture. The work in this category includes image quality measures [4], micro-texture analysis [3] and Difference of Gaussian (DOG)[105]. Different score-level fusion strategies are also applied to analyse its effect on spoofing attacks. Some of the representative work in this area are [99] and [106].

5.5 Experiments and Results

To identify the modalities in different HFR scenarios, three challenging databases have been selected i.e visual vs. NIR (HFB Face Database), low resolution (Web Camera) vs. high resolution (Digital Camera) (Biosecure Face Database)and visual vs sketch (CUHK Face Database). In testing one of the popular similarity measures, normalized correlation, is used to find the score between camera fingerprint and probe image noise residue.

In HFB database, the camera fingerprints for visual and NIR modalities are estimated by using 450 VIS-NIR images selected randomly for training while 1000 images of both modalities are used as probe in my experimental setup. All samples of VIS-NIR database are of the size 128 x 128 cropped using eye co-ordinates. For Biosecure face database, digital and web cameras SPN are calculated using 400 training images from both modalities while 200 digital and web-camera samples are used for testing.

In CUHK photo-sketch face database,88 photo-sketch pairs are used for training and 100 pairs are used as probe in testing phase. Again, all samples in Biosecure and CUHK face databases are the same size i.e. 128 x 128 as in HFB database. In my experimental setup , all training images from different modalities are selected randomly and are not included in the test set.

The test image MPN and the modality reference MPN are extracted from the green channel and the wavelet based de-noising filter is used. I follow the same experimental setup of Lukas et. al [94] approach by using green channel for all modalities. But it is worth to mention here that I can learn modality finger print for each channel and study its effect on error rates. One can also combine the different channel modality fingerprints using fusion techniques. My proposed UM based MPN produces attractive results in the identification of the modalities in different HFR setup. The error rates in HFB database are quite low and validate my notion that preprocessing based on unsharp-masking helps to achieve lower error rates i.e. 25% and 8.33% reduction in visual modality for basic and phase MPN respectively. Similarly, the error rates on CUHK database for photo and sketch modalities are quite promising and in phase MPN results, I obtain almost 43% reduction in error rate when UM is applied. For Biosecure database, the results for digital camera are very promising and unsharp masking again helping to reduce error rate while for webcam error rates are little bit higher. Table 5.1 presents the results on three HFR databases using my proposed MPN methods namely unsharp masking basic modality pattern noise method (UM-B-MPN) with Lukas et. al [94] approach (B-MPN) and unsharp masking phase modality pattern noise method (UM-Ph-MPN) with Kang et al. [95] method (Ph-MPN).

This study is the first approach to tackle modality identification problem and direct comparison with other approaches handling this problem is not possible. To carry out a fair comparison, I have selected two recent approaches from liveliness / anti-spoofing approaches [4] and [3] to compare with my proposed MPN method.

Figure 5-5 provides a bar chart comparison of error rates among my proposed MPN methods namely unsharp masking basic modality pattern noise method (UM-

Table 5.1: Error Rates on different HFR Databases

		Error Rate %			
HFR Database	Modality	B-MPN	UM-B-MPN	Ph-MPN	UM-Ph-MPN
HFB	VIS	0.4	0.3	0.12	0.11
	NIR	0	0	0.7	0.4
CUHK	Photo	0	0	5	5
	Sketch	13	11	16	14
Biosecure	Webcam	23.4	20	30.1	27.6
	Digcam	9.5	7	8.5	7.9

B-MPN) and unsharp masking phase modality pattern noise method (UM-Ph-MPN) with micro texture analysis method (MTA) by J. Määttä et. al [3] and image quality assessment measures (IQA) by Javier Galbally et. al. [4]. On HFB database, UM-B-MPN and MTA have zero percent error. MTA has quite outstanding results on Biosecure database while on CUHK, IQA and UM-B-MPN error rates are quite close to each other.

5.6 Discussion

HFR can have potentially various applications including the following three areas: surveillance, forensics and security. Some of the real-world application of these areas are Airports, Crime Investigation and Banks etc. To reflect these scenarios, since all the HFR databases are collected separately and managed to deal only with one type of modalities so researchers in this area have to work around this constraint. In my experimental setup during the decision process, I ended up performing binary classification as databases deal with only e.g visual vs. NIR. For more challenging task, I have combined the modalities of HFB and Biosecure databases samples except CUHK (photo-sketch) database as it has only 88 samples available for training. Table 5.2 represents the results of this scenario when I have tested 1 vs 4 e.g. visual vs (NIR, Digitalcam , Webcam). I train with 400 images of four modalities of two databases and test 200 samples from each with learned four fingerprints to get the correct error rate. The error rates in this scenario are very similar to previously mentioned results.

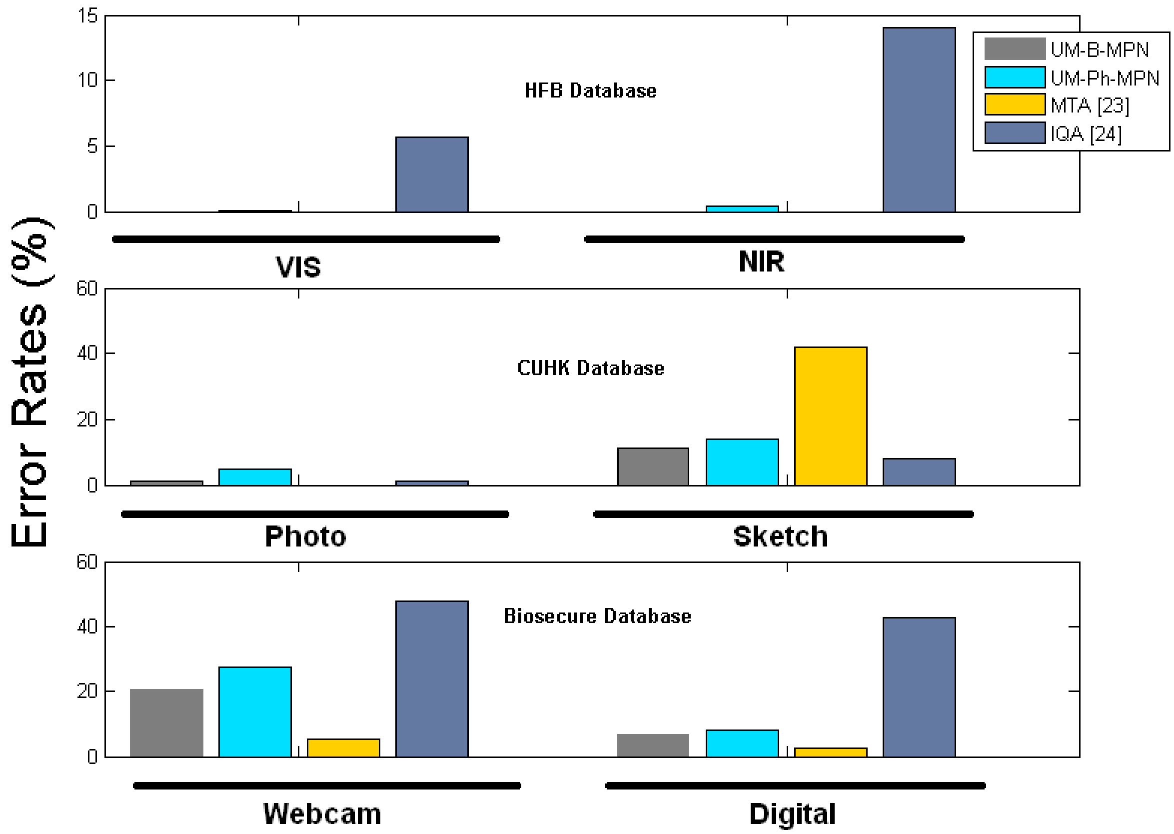

Figure 5-5: Bar chart comparison of error rates among proposed UM-B-MPN and UM-Ph-MPN methods vs. MTA [3] and IQA [4].

This is to be noted for MPN based fingerprint retrieval of different modalities that lower number of training samples and low resolution of images result in learning of weak reference MPN of modalities. But, my method tackles those issues by learning fingerprint with lower number of samples and low resolution of images to validate the efficacy of proposed approach.

5.7 Summary

Modality identification is an important component in fully heterogenous face recognition (HFR) systems since it is not possible to develop fully automated HFR systems without modality identification. In this paper, a novel method is proposed for modality identification which is inspired from sensor pattern noise (SPN) estimation based

Table 5.2: Error Rates when fingerprints of HFB and Biosecure tested in combination

		Error Rate %			
HFR Database	Modality	B-MPN	UM-B-MPN	Ph-MPN	UM-Ph-MPN
HFB	VIS	1	0.5	2.5	2.5
	NIR	0	0	0.8	0.6
Biosecure	Webcam	29.5	26.5	37	31.5
	Digcam	9.5	7	9.5	8.7

approaches in camera / source identification. The results of my proposed Unsharp Masking (UM) based Modality Pattern Noise (MPN) estimation brings an exciting opportunity for researchers in HFR domain to solve this problem. The proposed system has been evaluated using three challenging benchmarks of intermodality face matching: Biosecure (Low vs High) , CUHK (VIS vs Sketch) and HFB (VIS vs NIR). The proposed technique has produced outstanding results.

Chapter 6

Bagging based PLDA using Histogram of Gradients features for Heterogeneous Face Recognition

6.1 Introduction

Chapters 3 and 4 introduce probabilistic solution to the heterogeneous face recognition problem. Probabilistic discriminant analysis (PLDA) is the first ever probabilistic approach in HFR research domain. It provides theoretical framework for applying concepts of probability in face recognition problem of different modalities. It produces very good results on CASIA HFB and Biosecure databases comparing to the state-of-the art methods [75, 46, 81, 82, 83]. Probabilistic methods have far greater advantage over other statistical approaches e.g. careful modelling of noise , marginalization, coherent way of model comparison, deferment of final decision in case of big uncertainty etc. In proposed method, I manage to produce outstanding, comparable results merely using intensity features globally on VIS-NIR and Webcam-DigiCam databases. Experiments on PLDA have been carried out according to protocol setup of [107] which considers overlapping train and test sets.

Tied Factor analysis (TFA) is the specialized Bayesian based approach tackling small sample size (SSS) problem in homogeneous FR. I extend TFA to heterogeneous FR case where it can handle more than modalities. There are very few approaches [48] that are dealing SSS problem in HFR domain which is actually a real world scenario. The major disadvantage of TFA approach lies in producing biased results as it only deals with one set of gallery and probe images for VIS and NIR modalities respectively. I extensively test the TFA approach for heterogeneous face databases with cross validation leave-one-out strategy to report un-biased results. Further to improve recognition accuracies, I combine the probabilities acquired from individual heterogeneous face pairs. TFA manages to produce outstanding results on protocol I and II when compared with [75, 46, 81, 82]. It is worth to mention that it utilizes intensity based features only. I extend its experimental setup to a challenging third protocol setup termed as Protocol III which is a completely disjoint set and it reports better results compared to [46, 82, 83, 84]. There is a room of improvement in recognition accuracy for protocol III.

Following are the main contributions in this chapter

- Bagging based probabilistic linear discriminant analysis (PLDA) is an extension of PLDA method used in Chapter 2 for heterogeneous face recognition. PLDA can report biased results using subsets of those images that produced good recognition rates. Results of PLDA are also misleading as it uses overlapping train and test sets. Histogram of gradient descriptors (HOG) are used due to their already established importance in intermodality face matching. In this chapter, a bagging based PLDA is proposed to evaluate the challenging database HFB exhaustively. Bagging based PLDA recognition rates outperform all the state-of-the-art methods when using only HOG features. Results depict the strength and utilization of HOG features application in HFR domain when applied with proposed methods in this thesis.

This chapter explores the efficacy of Bagging based PLDA and TFA in heterogeneous face matching in variety of experimental setups i.e. global face matching , local face matching, global and local bagging PLDA , global and local bagging TFA.

The next section 6.2 discusses the related work in HFR domain. Then, I compare in section 6.3 the subtle differences between TFA and PLDA and my contributions. The next section 6.4 provides detailed experimental results on PLDA and TFA in different parameters setup and finally , I discuss the results in section 6.5 and provide concluding remarks in section 6.6.

6.2 Related Work

In this section, I discuss about the current existing techniques in heterogeneous face recognition dealing different modalities e.g. VIS vs NIR, Photo vs Sketch , VIS vs thermal etc. The work in this domain can be broadly categorized into modality invariant feature extraction, common subspace methods and face synthesis methods.

6.2.1 Modality Invariant Feature Extraction

The modality invariant features represent the heterogeneous modalities e.g. visual and near infrared in a common feature subspace which is insensitive to modalities of images. Some of the representative work in this area are already discussed in chapters 3 and 4 which include [43, 44, 45, 26, 79]. Most recently in [83], J. Zhu et. al. proposed a simple feature representation by applying log-DOG filter in conjunction with local encoding mechanism and uniform feature normalization to reduce the feature gap between heterogeneous face images. Their method utilizes transduction approach to learn discriminative model for classification. In [108], a learning based descriptor is proposed for feature extraction stage which transforms the heterogeneous face pairs into encoded versions where the modality gap is reduced. Their descriptor enhances the correlation between encoded images of same subject to improve the recognition performance.

6.2.2 Common Subspace Methods

The subspace methods try to seek a new subspace where the difference of modality is minimized. Some state-of-the-art methods [46, 75, 48] in this area are already given in chapters 3 and 4. A canonical correlation analysis (CCA) is used as a correlation mechanism to learn relationship between NIR and VIS faces in [109].

6.2.3 Face Synthesis Methods

The face synthesis analysis convert a query image's modality into gallery's modality by synthesising a pseudo-image for matching. [40, 41, 42, 77, 78] are analysis by synthesis methods already explained in chapters 3 and 4. Applying face analogy by Wang et al. in [110], they converted images from one modality type to another. They subsequently compared synthesized query images to gallery set as a patch of image has nearly the same similarity as its neighboring patches in VIS and the corresponding NIR domains. In [111], NIR image is transformed to produce synthesize VIS image along with pose rectification.

6.2.4 Histogram of Gradient (HOG) Descriptor Based Methods

Histogram of gradient descriptors are first utilized in human or object detection by N. Dalal et. al [112]. In calculating the HOG features, an image is firstly decomposed into equal small squared cells and then histogram of oriented gradients are computed in each cell. Finally, normalization is done on the resultant histograms via block-wise pattern so making descriptor of each cell. HOG feature descriptors are also employed in heterogeneous face recognition [44] but in combination with LBP feature descriptor. Some other state-of-the-art methods also compare their proposed feature descriptors with HOG [108, 83, 26].

Most recently, researchers from MIT have developed HOG goggles [113] to perceive the visual world as HOG based object detector. With context to its success in object detection, I employ the HOG feature solely on heterogeneous face database with my proposed methods in chapter 2 and 3 to find out the strength and performance of this feature descriptor. HOG features encode local shape information to capture spatial information from the small squared cells.

Figure 6-1 and Figure 6-2 represent HOG features visual display of selected VIS and NIR examples from HFB face database on different cell sizes. The HOG features on both modalities quite resemble to each other and is a good choice to opt as selected feature representation.

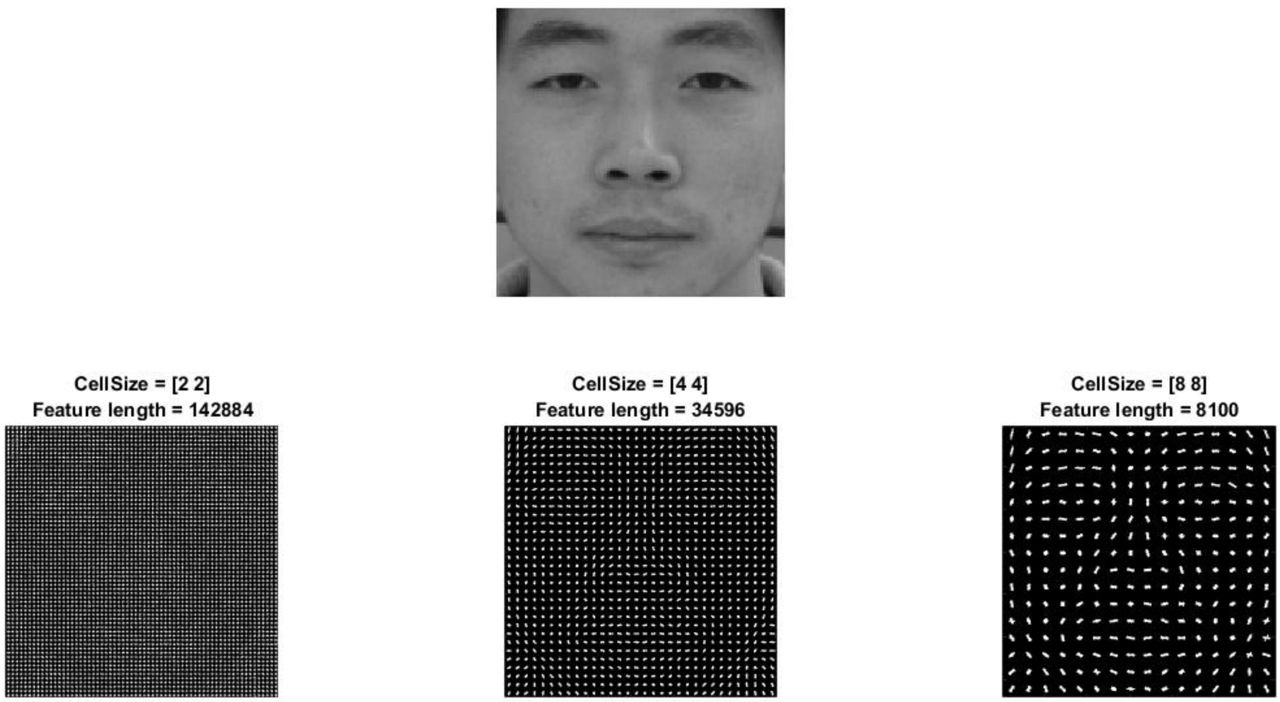

Figure 6-1: Representation of HOG feature descriptor of visual modality of HFB database

6.3 Comparison between PLDA and TFA

The subtle difference between PLDA [2] and TFA [32] algorithms lies in the formulation of the noise component.

PLDA model is of the form

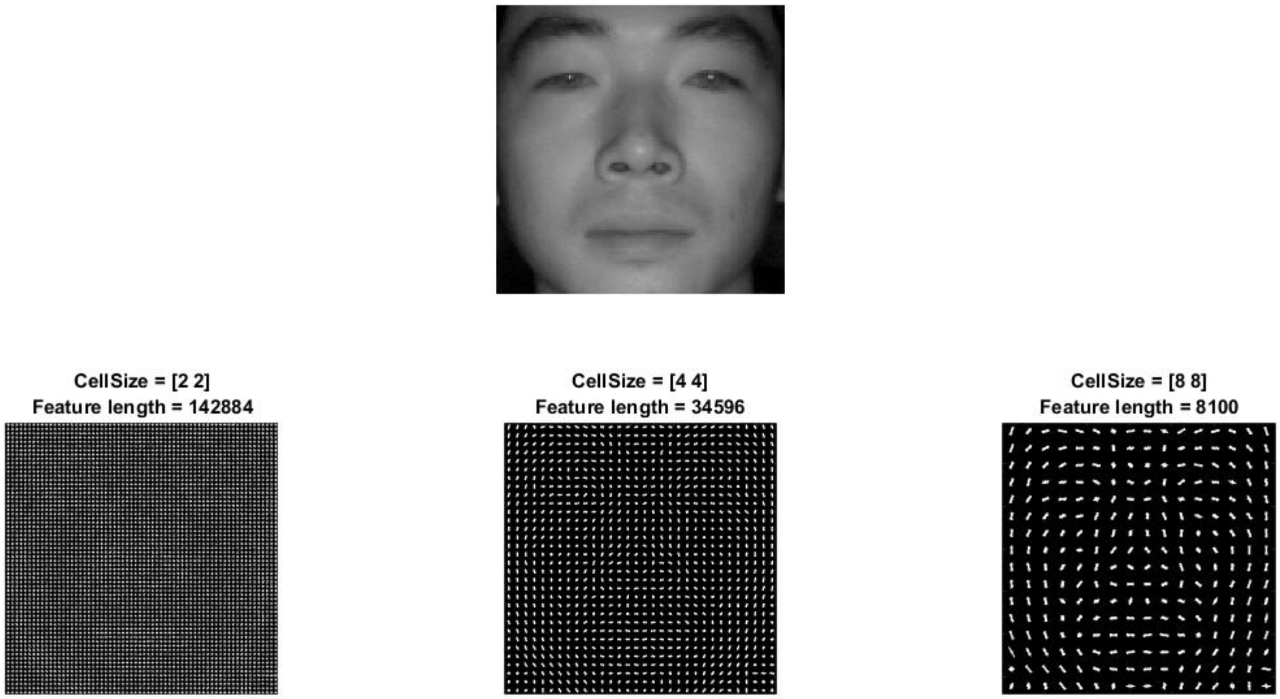

Figure 6-2: Representation of HOG feature descriptor of near infrared modality (NIR) of HFB database

$$x_{ij} = \mu + Fh_i + Gw_{ij} + \epsilon_{ij} \tag{6.1}$$

TFA model is of the form

$$\mathrm{x}_{ijk} = \mathrm{F}_k \mathrm{h}_i + \mathrm{m}_k + \epsilon_{ijk} \tag{6.2}$$

In PLDA equation 6.1, the extra noise term Gw_{ij} represents the variance in images of the same person while there is no such provision in TFA algorithm. This extra noise term adds complexity to the PLDA algorithm but definitely brings quite favorable characteristics as it re-formulates famous LDA method in probabilistic way. The PLDA method has signal component represented by $\mu + Fh_i$ describing intra-class variation while noise component represented by $Gw_{ij} + \epsilon_{ij}$ describing inter-class variation. The details of learning parameters of both algorithms is already explained in chapters 2 and 3.

6.4 Experiments

The robustness of proposed PLDA and proposed TFA algorithms for intermodality face matching are tested on two different types of intermodality scenarios i.e Visual vs. NIR and Low resolution (Web Camera) vs. High resolution (Digital Camera). As results are already reported on intensity and LBP features in my previous chapters, HOG feature descriptors are employed and rank one recognition rates are mentioned.

The extensive experiments are carried out on VIS-NIR HFB face database [49]. The HFB contains 202 individuals with total 5097 images out of which 2095 and 3002 images are of visual and NIR modality, respectively. In order to test the validity of my proposed methods in heterogeneous face recognition problem, I use the completely disjoint training and test sets. I adopt the same protocols setup of [83, 26] where 100 individuals selected as training and 100 individuals selected as the test images involving gallery and probe images. To add more complexity to the testing scenario, I randomly select train and test sets containing only single image of modality i.e. VIS and NIR.

Image pre-processing step used in [22] are also applied in my experiments before extracting HOG feature descriptors. I use 8x8 cell size and 2x2 block size of HOG parameter settings in my experiments. For global approach, HOG feature size is 8100D. I divide the image into 16x16 overlapping blocks of original image size of 128x128. In this way, I have total 64 patches of each image for my proposed methods local approach.

Figure 6-3 presents the example image and its patch based image of HFB face database.

Table 6.1 presents the rank one recognition rates on HFB database using same protocol mentioned above by applying HOG feature descriptors globally and locally. It is worth to mention that I use only single instance from each modalities images to test my proposed methods in real world scenarios. The feature size of HOG for global approach and local approach are 8100D and 36D. I also concatenate the 64 patch local features to make one long feature for my global approaches utilizing local

Figure 6-3: Original image and its local patch image example from HFB database

patch HOG features. The dimension of this feature vector is 2304D i.e. 64x36D

Table 6.1: Results using my proposed PLDA and TFA based on global and local HOG feature descriptors on HFB Face Database

	Recognition Rates (%)		
Method	Global HOG	Global HOG local patch data	Local Patch HOG
Bagging TFA [114]	47	32	65
PLDA for HFR [115]	72	62	76

It is evident from the Table 6.1 that proposed PLDA method has produced better results as compared to my proposed TFA method in all three different experiment settings. Definitely, PLDA's extra noise component help in capturing image fine details including shape and texture. Experiments also validate the superiority of local based approaches over global approaches as in case of TFA, the rank one recognition rate has 38% rise. The score matrices of global PLDA and TFA are represented in Figure 6-4. PLDA supremacy over TFA is further asserted by ROC and CMC curves in Figure 6-5.

6.5 Discussion

Table 6.2 compares the recognition rates on HFB using my proposed PLDA for HFR and Bagging TFA with other methods.

My proposed methods PLDA and TFA report better results on HOG based fea-

Table 6.2: Comparison of the recognition rates using my proposed PLDA and TFA based on HOG feature descriptors CASIA HFB

Method	Recognition Rates (%) Only on HOG Features
PCA	5.9
FDA [116]	53.5
PCA+CCA [109]	49.67
CSR [82]	10.17
tPCA [83]	7.33
TCA [117]	21.61
SDA [84]	51.94
THFM [83]	62
Bagging based TFA	65
Bagging based PLDA	**76**

tures and outperform all the stat-of-art heterogeneous face methods which reflect my proposed approaches effectiveness on other approaches. Further, it validates the HOG feature superiority over other computationally complex features e.g. discriminant face descriptor (DFD) [26], common feature discriminant analysis (CFDA) [108] etc. HOG feature extraction process is computationally cheaper than its counterparts.

6.6 Summary

In this chapter, I compare the efficacy of PLDA approach vs. TFA approach based on posterior probabilities in the intermodality face matching problem. PLDA shows some consistent results on visual vs near-infrared and digital-camera vs web-camera. It is also to be noted that many heterogenous face matching approaches use variety of statistical learning approaches but PLDA + TFA, a generative probabilistic approaches are applied first time in this domain. Both methods are experimented in more challenging case of heterogeneous face matching when there exists only small training samples. To report fair comparison between proposed PLDA and TFA, different evaluation settings based on global and local HOG feature descriptors are applied.

It is evident that the proposed PLDA method not only improves the results compared to proposed TFA model but also effectively outnumber other state-of-the-art approaches with good margin by just using HOG features.

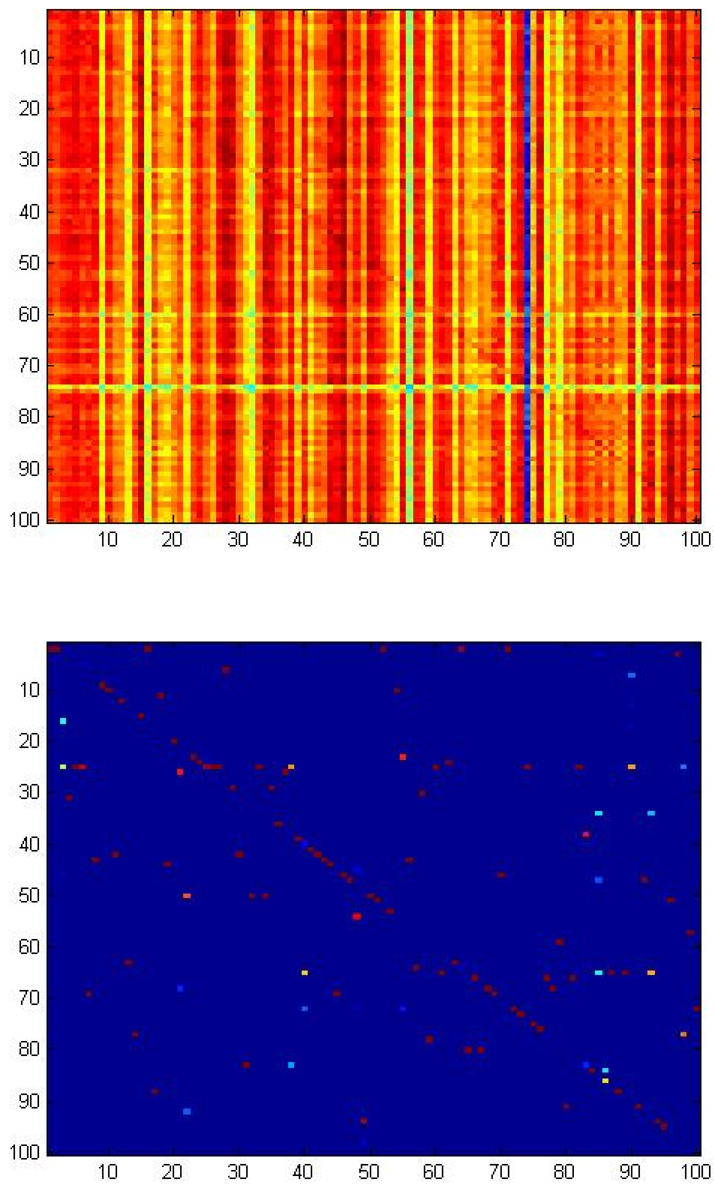

Figure 6-4: Score matrices of global PLDA and TFA

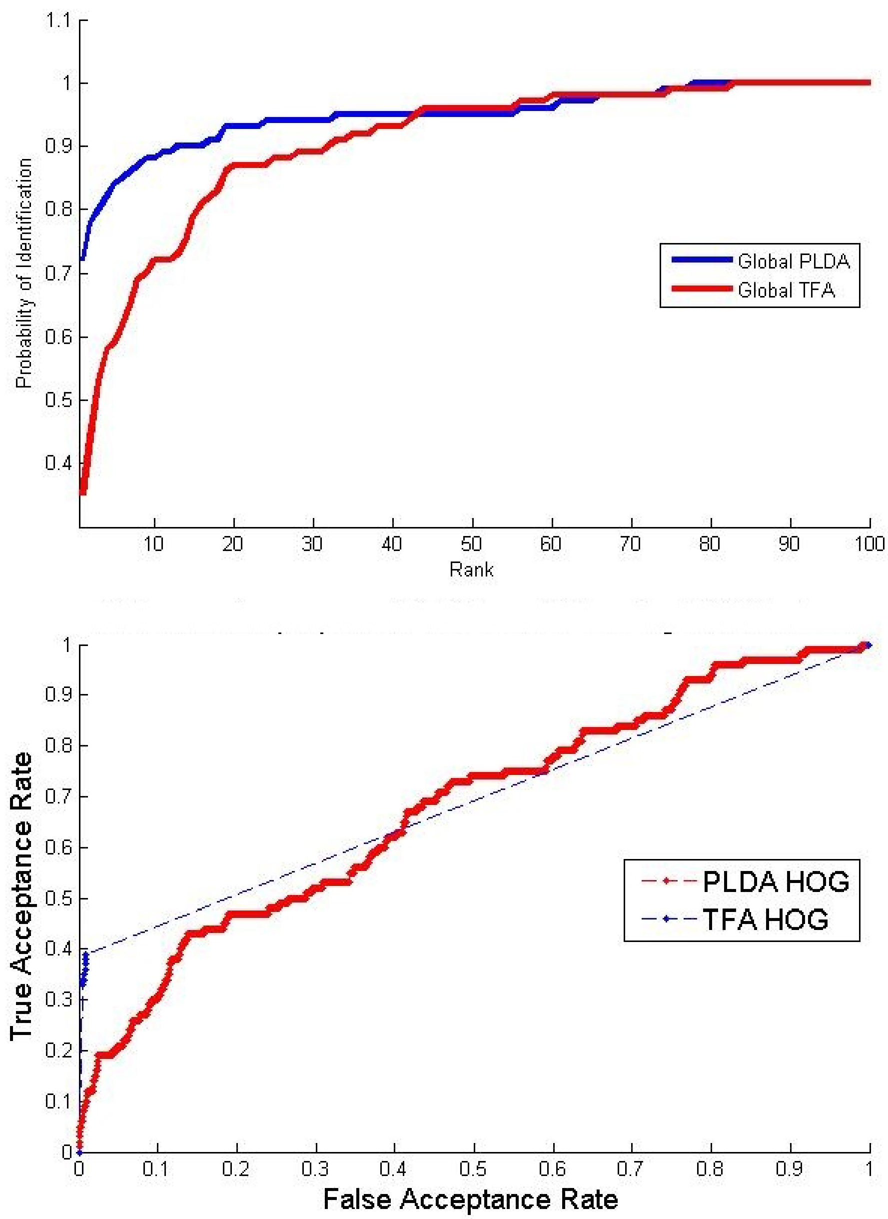

Figure 6-5: CMC an ROC curves for proposed global based PLDA and TFA

Chapter 7

Conclusions and Future Work

This thesis studied the problems of homogeneous and heterogeneous face matching. The main contributions are made towards representing faces in different modalities with generic representation using latent identity variable. This generative model using posterior probabilities laid the theoretical foundation of first ever probabilistic framework. Initial efforts to solve intermodality problem is done by utilizing intensity features in real world scenario i.e. when single image of person exists aka small sample space. Identifying types of modalities in heterogeneous face recognition is another primary contribution.

7.1 Conclusions

This thesis covered range of problems in homogeneous as well as heterogeneous face recognition. The probabilistic framework is proposed to handle the uncertain nature of face manifolds and bring favorable characteristics towards face matching. Modality identification module is the linchpin to the automation of heterogenous face recognition. The results are important to the research community to assist in range of problems in heterogeneous face recognition that are of interest in many biometric applications.

The multi-scale LBP feature descriptors that results in curse of dimensionality is a major barrier for multi-scale approaches. The forward feature selection study is the first step towards identifying the optimal feature subset for correct and accurate representation of an image. The initial results supports the notion that even after first few runs of feature concatenation result in reduction of variance among recognition rates of different scales.

Although, research on intermodality face recognition has reached to maturity level but in reality as compared to homogeneous face recognition, there is lot to achieve before embarking any practical applications of this approach. The CASIA face databases are not a representative database as it does not cover entire world demographics. No statistics and standards are available for heterogeneous face recognition as compared to homogeneous face recognition where Face Recognition Vendor Test (FRVT) 2002 and the Face Recognition Grand Challenge (FRGC) present to validate and authenticate new methods.

7.2 Future Work

Although contributions presented in this thesis covered range of problems in homogeneous and heterogeneous face recognition but there are still many challenges remain that has to be given due consideration.

The possible future direction of PLDA and TFA methods will be to introduce Bayesian regularization and to apply it in different variants of NIR i.e. SWIR, MWIR etc. and source identification.

Although, I have received comparable and competitive results but considerable gain can be attained by using local PLDA models i.e. separate PLDA model may be built on each manual or automatic annotations of the face.

The proposed bagging TFA method can be extended to local patch matching approach by learning local TFA models in combination with available powerful feature descriptors e.g. LBP, HOG etc. Further, original TFA has established record in large pose perturbation so if any future heterogeneous database brings variation in pose then the proposed method can be considered an automatic choice.

In future, other feature selection strategies may be employed e.g. heuristic or greedy approaches to study the effect of feature selection on different scales or radii of multi-scale approaches. The probabilistic approaches has to be researched in depth due to its far greater advantages over other statistical approaches.

Absence of modality identification module will definitely hamper in bringing the automated face recognition systems. Its importance is similar as head pose estimation module, facial expression identification module, camera identification module, liveliness detection module etc. But in order to fully research this grey area, an erection of standard heterogeneous face database is required covering all type of modalities meaning gallery images of individuals should be higher resolution and presence of other modalities images e.g. NIR, thermal, sketch, low resolution may be web-cam or video etc. should be there to comprehensively test and validate any new state-of-the-art methods.

Deep learning methods has numerous advantages over major statistical approaches which can be labelled as shallow machine learning. There is a flurry of research in other areas but there are very little or minimal effort in heterogeneous face recognition to identify the strengths and weaknesses of deep learning algorithms.

Bibliography

[1] M. El-Abed, C. Charrier, and C. Rosenberger, "Evaluation of biometric systems," *INTECH Open Access Publisher*, 2012.

[2] P. Li, Y. Fu, U. Mohammed, J. H. Elder, and S.J.D. Prince, "Probabilistic models for inference about identity," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 34, no. 1, pp. 144–157, 2012.

[3] J. Määttä, A. Hadid, and M. Pietikainen, "Face spoofing detection from single images using micro-texture analysis," *IEEE International joint conference on Biometrics*, pp. 1–7, 2011.

[4] J. Galbally, S. Marcel, and J. Fierrez, "Image quality assessment for fake biometric detection: Application to iris, fingerprint, and face recognition," *IEEE Transactions on Image Processing*, vol. 23, no. 2, pp. 710–724, 2014.

[5] S. Prabhakar, S. Pankanti, and A. K. Jain, "Biometric recognition: Security and privacy concerns," *IEEE Security and Privacy*, pp. 33–42, 2003.

[6] K. P. Tripathi, "A comparative study of biometric technologies with reference to human interface," *International Journal of Computer Applications*, vol. 14, no. 5, pp. 10–15, 2011.

[7] K. Delac and M. Grgic, "A survey of biometric recognition methods," *IEEE 46th International Symposium in Electronics in Marine*, pp. 184–193, 2004.

[8] T. Huang, Z. Xiong, and Z. Zhang, "Face recognition applications," *Handbook of Face Recognition*, pp. 617–638, 2011.

[9] A. K. Jain and S. Z. Li, "Handbook of face recognition," *New York : Springer*, vol. 1, 2005.

[10] B. Abidi and M. A. Abidi, "Face biometrics for personal identification: multi-sensory multi-modal systems," *Springer Science and Business Media*, 2007.

[11] P. Jonathon Phillips, Hyeonjoon Moon, Syed A. Rizvi, and Patrick J. Rauss, "The feret evaluation methodology for face-recognition algorithms," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 22, no. 10, pp. 1090–1104, 2000.

[12] S. Ferdinando and A. Harter, "Parameterisation of a stochastic model for human face identification," *2nd IEEE Workshop on Applications of Computer Vision*, 1994.

[13] Aleix Martínez and Robert Benavente, "The ar face database," , no. 24, 1998.

[14] T. Sim, S. Baker, and M. Bsat, "The cmu pose, illumination, and expression (pie) database," *IEEE International Conference on Automatic Face and Gesture Recognition*, p. 46âĂŞ51, 2002.

[15] A. S. Georghiades, P. N. Belhumeur, and D. J. Kriegman, "From few to many: Illumination cone models for face recognition under variable lighting and pose," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 23, no. 6, pp. 643–660, 2001.

[16] W. Zha, R. Chellappa, P.J. Phillips, and A. Rosenfeld, "Face recognition: A literature survey," *ACM Computing Surveys*, vol. 35, no. 4, pp. 399–458, 2003.

[17] T. Ahonen, A. Hadid, and M. Pietikainen, "Face description with local binary patterns : Application to face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 28, no. 12, pp. 2037–2041, 2006.

[18] C. Liu and H. Wechsler, "Gabor feature based classification using the enhanced fisher linear discriminant model for face recognition," *IEEE Transactions on Image processing*, vol. 11, no. 4, pp. 467–476, 2002.

[19] L. Shen and L. Bai, "Mutualboost learning for selecting gabor features for face recognition," *Pattern Recognition Letters*, vol. 27, no. 15, pp. 1758–1767, 2006.

[20] C. Liu and H. Wechsler, "Independent component analysis of gabor features for face recognition," *IEEE Transactions on Neural Networks*, vol. 14, no. 4, pp. 919–928, 2003.

[21] W. Zhang, S. Shan, W. Gao, X. Chen, and H. Zhang, "Local gabor binary pattern histogram sequence (lgbphs): A novel non-statistical model for face representation and recognition," *IEEE International Conference on Computer Vision, ICCV*, pp. 786–791, 2005.

[22] X. Tan and B. Triggs, "Enhanced local texture feature sets for face recognition under difficult lighting conditions," *IEEE Transactions on Image Processing*, vol. 19, no. 6, pp. 1635–1650, 2010.

[23] S. U. Hussain, T. Napoléon, and F. Jurie, "Face recognition using local quantized patterns," *British Machive Vision Conference*, pp. 1–11, 2012.

[24] S. Liao, X. Zhu, Z. Lei, L. Zhang, and S.Z. Li, "Learning multi-scale block local binary patterns for face recognition," *International Conference on Advances in Biometrics (ICB)*, pp. 828–837, 2007.

[25] T. Ojala, M. Pietikainen, and T. Maenpaa, "Multiresolution gray-scale and rotation invariant texture classification with local binary patterns," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 24, no. 7, pp. 971–987, 2002.

[26] Z. Lei, M. Pietikaĺ inen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, pp. 289–302, February 2014.

[27] J. Wright, A. Y. Yang, A. Ganesh, S. S. Sastry, and Y. Ma, "Robust face recognition via sparse representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 2, pp. 210–227, 2009.

[28] I. Naseem, R. Togneri, and M. Bennamoun, "Linear regression for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, no. 11, pp. 2106–2112, 2010.

[29] J. Yang, D. Zhang, A. F. Frangi, and J. Y. Yang, "Two dimensional pca: a new approach to appearance-based face representation and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 26, no. 1, pp. 131–137, 2004.

[30] L. Zhang, M. Yang, and X. Feng, "Sparse representation or collaborative representation: Which helps face recognition?," *IEEE International Conference on Computer Vision*, 2011.

[31] P. Li and S. J. Prince, "Probabilistic methods for face registration and recognition," *Advances in Face Image Analysis: Techniques and Technologies*, 2010.

[32] S. J. Prince, J. Warrell, J. H. Elder, and F. M. Felisberti, "Tied factor analysis for face recognition across large pose differences," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, pp. 970–984, June 2008.

[33] P. N. Belhumeur, J. P. Hespanha, and D. J. Kriegman, "Eigenfaces vs fisherfaces: recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[34] M. Turk and A. Pentland, "Eigenfaces for recognition," *Journal of Cognitive Neuroscience*, vol. 3, no. 1, pp. 71–86, 1991.

[35] B. Moghaddam, W. Wahid, and A. Pentland, "Beyond eigenfaces: Probabilistic matching for face recognition," *IEEE International Conference on Automatic Face and Gesture Recognition*, pp. 30–35, 1998.

[36] B. Moghaddam and A. Pentland, "Probabilistic visual learning for object representation," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 696–710, 1997.

[37] B. Moghaddam, T. Jebara, and A. Pentland, "Bayesian face recognition," *Pattern Recognition*, vol. 33, no. 11, pp. 1771–1782, 2000.

[38] C. Liu and H. Wechsler, "Probabilistic reasoning models for face recognition," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 827–832, 1998.

[39] S. Ioffe, "Probabilistic linear discriminant analysis," *European Conference on Computer Vision*, pp. 531–542, 2006.

[40] T. Xiaoou and X. Wang, "Face sketch recognition.," *IEEE Transaction, Circuits and Systems for Video Technology*, vol. 14, pp. 50–57, 2004.

[41] Q. Liu, X. Tang, H. Jin, H. Lu, and S. Ma, "A non-linear approach for face sketch synthesis and recognition," *IEEE Conference on Computer Vision Pattern Recognition*, vol. 1, pp. 1005–1010, 2005.

[42] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, pp. 1955–1967, 2009.

[43] S. Liao, D. Yi nad Z. Lei, R. Qin, and S. Z. Li, "Heterogeneous face recognition from local structures of normalized appearance," *Springer, In Advances in Biometrics*, pp. 209–218, 2009.

[44] B. Klare and A. K. Jain, "Heterogeneous face recognition: Matching nir to visible light images," *International Conference on Pattern Recognition*, pp. 1513–1516, 2010.

[45] B. Klare and Z. Li andA. K. Jain, "Matching forensic sketches to mug shot photos," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 33, pp. 639–646, 2011.

[46] X. Huang, Z. Lei, X. Wang, and S. Z. Li, "Regularized discriminative spectral regression method for heterogeneous face matching," *IEEE Transactions on Image Processing*, vol. 22, pp. 1513–1516, 2013.

[47] Z. Lei, S. Liao, A. K. Jain, and S. Z. Li, "Coupled discriminant analysis for heterogeneous face recognition," *IEEE Transactions on Information Forensics and Security*, vol. 7, no. 6, pp. 1707–1716, 2012.

[48] B. F. Klare and A. K. Jain, "Heterogeneous face recognition using kernel prototype similarities," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 35, pp. 1410–1422, June 2013.

[49] S. Z. Li, Z. Lei, and M. Ao, "The hfb face database for heterogeneous face biometrics research," pp. 1–8, 2009.

[50] J. O. Garcia, J. Fierrez, F. A. Fernandez, J. Galbally, M. R. Freire, J. Gonzalez-Rodriguez, C. Garcia-Mateo, J. L. Alba-Castro, E. Gonzalez-Agulla, E. Otero-Muras, S. Garcia-Salicetti, L. Allano, B. Ly-Van, B. Dorizzi, J. Kittler, T. Bourlai, N. Poh, F. Deravi, M. Ng, M. Fairhurst, J. Hennebert, A. Humm, M. Tistarelli, L. Brodo, J. Richiardi, A. Drygajlo, H. Ganster, F. M. Sukno, S. K. Pavani, A. Frangi, L. Akarun, and A. Savran, "The multiscenario multi-environment biosecure multimodal database (bmdb)," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 32, pp. 1097–1111, April 2010.

[51] X. Wang and X. Tang, "Face photo-sketch synthesis and recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 31, no. 11, pp. 1955–1967, 2009.

[52] B. Zhang, Y. Gao, S. Zhao, and J. Liu, "Local derivative pattern versus local binary pattern: Face recognition with high-order local pattern descriptor," *IEEE Transaction on Image Processing*, vol. 19, no. 2, pp. 533–544, 2010.

[53] R. Chellappa, C. Wilson, and S. Sirohey, "Human and machine recognition of faces: A survey," *Proceedings of the IEEE*, vol. 83, no. 5, pp. 705–740, 1995.

[54] A. K. Jain and S. Z. Li, "Handbook of face recognition," *Springer-Verlag*, 2004.

[55] K.W. Bowyer, K. Chang, and P. Flynn, "A survey of approaches and challenges in 3d and multi-modal 3d + 2d face recognition," *Computer Vision and Image Understanding*, vol. 101, pp. 1–15, 2006.

[56] J. Zou, Q. Ji, and G. Nagy, "A comparative study of local matching approach for face recognition," *IEEE Transaction on Image Processing*, vol. 16, no. 10, pp. 2617–2628, 2007.

[57] S. H. Gao, I. Tsang, and L. Chia, "Kernel sparse representation for image classification and face recognition," *IEEE European Conference on Computer Vision*, 2010.

[58] B. Cheng, J. Yang, S. Yan, Y. Fu, and T.Huang, "Learning with l1-graph for image analysis," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 4, pp. 858–866, 2010.

[59] J. Wright, K. Yu, Y. Gong, and T. Huang, "Sparse representation for computer vision and pattern recognition," *Special Issue on Applications of compressive sensing and sparse representation*, vol. 98, no. 6, pp. 1031–1044, 2010.

[60] Pietikäinen M. and Ojala T., "Texture analysis in industrial applications," *Image Technology - Advances in Image Processing, Multimedia and Machine Vision, (Ed. J.L.C. Sanz), Springer-Verlag,*, pp. 337–359, 1996.

[61] M. A. Tahir and A. Bouridane, "Face recognition using kernel collaborative representation and multiscale local binary patterns," *IET Conference on Image Processing*, 2012.

[62] C. H. Chan, J. Kittler, and K. Messer, "Multi-scale local binary pattern histograms for face recognition," *International Conference on Bioinformatics*, 2007.

[63] J. Zhang, M. Marszalek, S. Lazebnik, and C. Schmid, "Local features and kernels for classification of texture and object categories: A comprehensive study," *International Journal of Computer Vision*, vol. 73, no. 2, pp. 213–238, 2007.

[64] Thomas Hofmann, Schölkopf Bernhard, and J. Smola Alexander, "Kernel methods in machine learning.," *The annals of statistics*, vol. 36, no. 3, pp. 1171–1220, 2008.

[65] J. Kittler, "Feature selection and extraction," *Handbook of pattern recognition and image processing*, pp. 59–83, 1986.

[66] R. Kohavi and G. H. John, "Wrappers for feature subset selection," *Artificial intelligence*, vol. 97, no. 1, pp. 273–324, 1997.

[67] X. Jiang, B. Mandal, and A. Kot, "Eigenfeature regularization and extraction in face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 30, no. 3, pp. 383–394, 2008.

[68] J. Lu, K.N. Plataniotis, A.N. Venetsanopoulos, and A. Kot, "Ensemble-based discriminant learning with boosting for face recognition," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 17, no. 1, pp. 166–178, 2006.

[69] Y. Saeys, I. Inza, and LarraÃśaga, "A review of feature selection techniques in bioinformatics," *Bioinformatics*, vol. 23, no. 19, pp. 2507–2517, 2007.

[70] T. Zhang, Y. Y. Tang, B. Fang, Z. Shang, and X. Liu, "Face recognition under varying illumination using gradientfaces," *IEEE Transaction on Image Processing*, vol. 18, no. 11, pp. 2599–2606, 2009.

[71] R. Fergus, P. Perona, and A. Zisserman, "Object class recognition by unsupervised scale invariant learning," *IEEE Conference on Computer Vision and Pattern Recognition*, pp. 264–271, 2003.

[72] M. Isard and A. Blake, "Condensation conditional density propagation for visual tracking," *International Journal of Computer Vision*, vol. 29, pp. 5–28, 1998.

[73] C. Rother, V. Kolmogorov, and A. Blake, "Grabcut: Interactive foreground extraction using iterated graph cuts," *ACM Transactions on Graphics (TOG)*, pp. 309–314, 2004.

[74] A. P. Dempster, N. M. Laird, and D. B. Rubin, "Maximum likelihood from incomplete data via the em algorithm," *Journal of the royal statistical society. Series B (methodological)*, vol. 39, pp. 1–38, 1977.

[75] D. Lin and X. Tang, "Inter-modality face recognition," *IEEE European Conference on Computer Vision*, pp. 13–26, 2006.

[76] S. Mika, G. Ratsch, J. Weston, B. Scholkopf, and K. R. Mullers, "Fisher discriminant analysis with kernels," *Proc. IEEE Neural Networks for Signal Processing*, pp. 41–48, 1999.

[77] Y. Ma, W. Xie, X. Chen, M. Liu, Y. Guo, J. Meng, and Q. Qin, "An analysis-by-synthesis method based on sparse representation for heterogeneous face biometrics," pp. 148–151, 2011.

[78] Z. Zhang, Y. Wang, Z. Zhang, and G. Zhang, "Face synthesis from near-infrared to visual light spectrum using quotient image and kernel-based multifactor analysis," pp. 1–4, 2011.

[79] J. Y. Zhu, W. S. Zheng, and J. H. Lai, "Logarithm gradient histogram: A general illumination invariant descriptor for face recognition," pp. 1–8, 2013.

[80] M. A. Tahir, J. Kittler, and F. Yan, "Inverse random under sampling for class imbalance problem and its application to multi-label classification.," *Science Direct, Pattern Recognition*, vol. 45, pp. 3738–3750, October 2012.

[81] Z. Lei, C. Zhou, D. Yi, S. Z. Li, and A. K. Jain, "An improved coupled spectral regression for heterogeneous face recognition," pp. 7–12, 2012.

[82] Z. Lei and S. Z. Li, "Coupled spectral regressoin for matching heterogeneous faces," pp. 1123–1128, 2009.

[83] J. Y. Zhu, W. S. Zheng, J. H. Lai, and S. Z. Li, "Matching nir face to vis face using transduction," *IEEE Transactions on Information Forensics and Security*, vol. 9, pp. 501–514, March 2014.

[84] D. Cai, X. He, and J. Han, "Semi-supervised discriminant analysis," pp. 1–7, 2007.

[85] Z. Lei, M. Pietikainen, and S. Z. Li, "Learning discriminant face descriptor," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 2, pp. 289–302, 2014.

[86] A. Lawgaly, F. Khelifi, and A. Bouridane, "Image sharpening for efficient source camera identification based on sensor pattern noise estimation," *Fourth IEEE International Conference In Emerging Security Technologies (EST)*, pp. 113–116, 2013.

[87] X. Cai, C. Wang, B. Xiao, X. Chen, Z. Lv, and Y. Shi, "Coupled latent least squares regression for heterogeneous face recognition," *IEEE International Conference in Image Processing (ICIP)*, pp. 2772–2776, 2013.

[88] Z. Li, D. Gong, Y. Qiao, and D. Tao, "Common feature discriminant analysis for matching infrared face images to optical face images," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2436–2445, 2014.

[89] J. Y. Zhu, W. S. Zheng, J. H. Lai, and S. Z. Li, "Matching nir face to vis face using transduction," *IEEE Transactions on Information Forensics and Security*, vol. 9, no. 3, pp. 501–514, 2014.

[90] D. Yi, Z. Lei, S. Liao, and S. Z. Li, "Shared representation learning for heterogeneous face recognition," *IEEE International Conference on Automatic Face and Gesture Recognition*, 2015.

[91] S. Z. Li, D. Yi, Z. Lei, and S. Liao, "The casia nir-vis 2.0 face database," *IEEE International Conference In Computer Vision and Pattern Recognition Workshops*, pp. 348–353, 2013.

[92] Y. H. Lee and S. Y. Park, "A study of convex/concave edges and edge-enhancing operators based on the laplacian," *IEEE Transactions on Circuits and Systems*, vol. 37, no. 7, pp. 940–946, 1990.

[93] Y. Sutcu, S. Bayram, H. T. Sencar, and N. Memon, "Improvements on sensor noise based source camera identification," *IEEE International Conference on Multimedia and Expo*, pp. 24–27, 2007.

[94] J. Lukas, J. Fridrich, and M. Goljan, "Digital camera identification from sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 1, pp. 205–214, 2006.

[95] X. Kang, Y. Li, Z. Qu, and J. Huang, "Enhancing source camera identification performance with a camera reference phase sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 7, pp. 393–402, 2012.

[96] C. T. Li, "Source camera identification using enhanced sensor pattern noise," *IEEE Transactions on Information Forensics and Security*, vol. 5, no. 2, pp. 280–287, 2010.

[97] M. Chen, J. Fridrich, M. Goljan, and J LukÃąs, "Determining image origin and integrity using sensor noise," *IEEE Transactions on Information Forensics and Security*, vol. 3, pp. 74–90, 2008.

[98] I. Pavlidis and P. Symosek, "The imaging issue in an automatic face/disguise detection system," *IEEE Workshop on Computer Vision Beyond the Visible Spectrum: Methods and Applications*, pp. 15–24, 2000.

[99] T. de Freitas Pereira, A. Anjos, J. M. De Martino, and S. Marcel, "Can face anti-spoofing countermeasures work in a real world scenario," *IEEE International conference on Biometrics (ICB)*, pp. 1–8, 2013.

[100] J. Komulainen, A. Hadid, M. Pietikainen, A. Anjos, and S. Marcel, "Complementary countermeasures for detecting scenic face spoofing attacks," *IEEE International conference on Biometrics (ICB)*, pp. 1–7, 2013.

[101] N. Kose and J. L. Dugelay, "Countermeasure for the protection of face recognition systems against mask attacks," *IEEE International Conference and Workshops on Automatic Face and Gesture Recognition*, pp. 1–6, 2013.

[102] G. Pan, L. Sun, Z. Wu, and S. Lao, "Eyeblink-based anti-spoofing in face recognition from a generic webcamera," *IEEE 11th Interntaional Conference on Computer Vision (ICCV)*, pp. 1–8, 2007.

[103] K. Kollreider, H. Fronthaler, and J. Bigun, "Non-intrusive liveness detection by face images," *Image and Vision Computing*, vol. 27, no. 3, pp. 233âĂŞ244, 2009.

[104] T. de Freitas Pereira, A. Anjos, J. M. de Martino, and S. Marcel, "Lbp-top based countermeasure against face spoofing attacks," *Computer Vision-ACCV Workshops*, pp. 121–132, 2012.

[105] Z. Zhang, J. Yan, S. Liu, Z. Lei, D. Yi, and S. Z. Li, "A face antispoofing database with diverse attacks," *IEEE International conference on Biometrics (ICB)*, pp. 26–31, 2012.

[106] I. Chingovska, A. Anjos, and S. Marcel, "Anti-spoofing in action: Joint operation with a verification system," *IEEE International on Computer Vision and Pattern Recognition Workshops (CVPRW)*, pp. 98–104, 2013.

[107] X. Huang, Z. Lei, M. Fan, X. Wang, and S. Z. Li, "Regularized discriminative spectral regression method for heterogeneous face matching," *IEEE Transactions on Image Processing*, vol. 22, pp. 353–362, January 2013.

[108] Z. Li, D. Gong, Y. Qiao, and D. Tao, "Common feature discriminant analysis for matching infrared face images to optical face images," *IEEE Transactions on Image Processing*, vol. 23, no. 6, pp. 2436–2445, 2014.

[109] D. Yi, R. Liu, R. Chu, Z. Lei, and S. Z. Li, "Face matching between near infrared and visible light images," *IEEE INTERNATIONAL CONFERENCE ON BIOMETRICS*, pp. 523–530, 2007.

[110] R. Wang, J. Yang, D. Yi, and S. Z. Li, "An analysis-by-synthesis method for heterogeneous face biometrics," *IEEE INTERNATIONAL CONFERENCE ON BIOMETRICS*, pp. 319–326, 2009.

[111] X. Pengfei, L. Huang, and C. Liu, "A method for heterogeneous face image synthesis," *IEEE INTERNATIONAL CONFERENCE ON BIOMETRICS*, pp. 1–6, 2012.

[112] N. Dalal and B. Triggs, "Histograms of oriented gradients for human detection," *IEEE Computer Vision and Pattern Recognition, CVPR*, pp. 886–893, 2005.

[113] C. Vondrick, A. Khosla, T. Malisiewicz, and A. Torralba, "Hoggles: Visualizing object detection features," *IEEE International Conference on Computer Vision (ICCV)*, pp. 1–8, 2013.

[114] M. K. Shaikh, M. A. Tahir, and A. Bouridane, "Tied factor analysis using bagging for heterogeneous face recognition," *IEEE, 5th European Workshop on Visual Information Processing (EUVIP)*, pp. 1–6, 2014.

[115] M. K. Shaikh, M. A. Tahir, and A. Bouridane, "Probabilistic linear discriminant analysis for intermodality face recognition," *IEEE International Conference on Acoustics, Speech and Signal Processing (ICASSP)*, pp. 509–513, 2014.

[116] P. Belhumeur, J. Hespanha, and D. Kriegman, "Eigenfaces vs. fisherfaces: Recognition using class specific linear projection," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 19, no. 7, pp. 711–720, 1997.

[117] B. Schölkopf, A. Smola, and K. Müller, "Nonlinear component analysis as a kernel eigenvalue problem," *Neural Computing*, vol. 10, no. 5, pp. 1299–1319, 1998.